

**URBVIZ: VISUAL ANALYSIS OF URBAN INDICES AND
GEOGRAPHICALLY AWARE CAUSALITY**

by

Micaela A. Melo

Submitted in partial fulfillment of the requirements
for the degree of Master of Computer Science

at

Dalhousie University
Halifax, Nova Scotia
December 2022

© Copyright by **Micaela A. Melo**, 2022

Table of Contents

List of Tables	iv
List of Figures	v
Abstract	vii
Acknowledgements	viii
Chapter 1 Introduction	1
1.1 Motivation	2
1.2 Contributions	2
1.3 Thesis Outline	3
Chapter 2 Related Work	4
2.1 Multi-attribute ranking visualization	4
2.2 Urban hotspot analysis	5
2.3 Causal relation visualizations	9
2.4 Visual urban analytics	12
2.5 Summary	13
Chapter 3 UrbViz	14
3.1 Datasets	14
3.2 Requirements	15
3.3 System	17
3.3.1 System architecture	17
3.3.2 Data processing pipelines	18
3.3.3 Visual Design	29
Chapter 4 Results and Evaluation	32
4.1 Pre-study questionnaire	32
4.2 Execution of tasks	33

4.3	Post-study questionnaire	36
4.4	Conclusion	38
Chapter 5	Conclusions and Future Work	39
5.1	Limitations and Future Work	40
Bibliography	41
Appendix A	Letter of Approval	46

List of Tables

3.1	Causal matrix for <i>total_distracted_driving</i> and <i>total_aggressive_driving</i> from traffic collisions dataset	28
-----	--	----

List of Figures

2.1	Interface of SmartAdP: interactive visual analytics system to solve billboard placements. A framework that integrates a novel visualization-driven data mining model enables advertising planners to effectively and efficiently formulate good candidate solutions.	5
2.2	Interface of SRVis: a spatial ranking visualization tool that supports efficient spatial multi-criteria decision-making processes.	6
2.3	Interface of Mirante: crime mapping visualization system that allows spatio-temporal analysis of crime patterns on a street-level scale using street-level heatmaps.	7
2.4	CrimAnalyzer System: Visual analytic tool that allows users to study the behaviour of crimes in specific city regions.	8
2.5	Visual Causality Analyst framework: reasoning framework that allows users to apply their expertise, verify and edit causal links, and collaborate with the causal discovery algorithm to identify a valid causal network. Its interface consists of both an interactive 2D graph view and a numerical presentation of salient statistical parameters, such as regression coefficients, p-values, and others.	10
2.6	Interface of Causality Explorer: A visualization tool that allows analysts to explore, validate, and apply causal relations in real-world decision-making scenarios. The tool provides an uncertainty-aware causal graph visualization for presenting a large set of causal relations inferred from high-dimensional data.	11
3.1	Rakings 0 and 1 in the Ranking View.	18
3.2	UrbViz contains three main views: (A) Ranking View, (B) Map View, and (C) Causality View. The Ranking View has 2 sub views: (A') is the menu ranking where the users select the datasets they want to work with and (A') presents two single-column rankings, one from the CIMD and the second from an urban dataset. The Map View shows a map with five different layers: hotspots, coldspots, relationship map, and pointwise locations of buildings and parks. The Causality View displays a causal network composed of nodes representing the indicators from Ranking 1, which are connected by links that denote causal influence.	19

3.3	Hotspots (Red hexagons) and Coldspots (Blue hexagons) shown in the Map View.	20
3.4	Fig. (Left) shows the significance map for the indicators chosen. Fig. (Right) shows the cluster map for the indicators chosen. .	23
3.5	Relationship Map layer shown in the Map View.	24
3.6	Legend (Color scale) of the Relationship Map.	25
3.7	Pointwise location of parks and buildings in the Map View. . .	26
3.8	A causal graph shown in the Causality View.	27
3.9	First 9 rows of Ranking 0 are highlighted in Ranking 1, and the visualizations of the map and causal graph are updated with the information of these areas.	29
4.1	Causal graph for the indicators of Ranking 1 for the entire municipality region.	34
4.2	Fig. (Left) shows the relationship map between 'Economic dependency scores' and 'ROW Maintenance, ROW lights, Traffic Inquiries'. Fig. (Right) shows the relationship map between 'Residential instability scores' and 'ROW Maintenance, ROW lights, Traffic Inquiries'	35
4.3	Fig. shows the causal graph of the entire municipality region .	36
4.4	Fig. shows the causal graph of the first 10 areas of Ranking 1 sorted by False Alarms.	37

Abstract

Municipalities and government entities collect massive amounts of daily data from urban activities and events for multiple purposes. Part of this information is used to base their decisions on geographically-targeted budgeting and resource allocation. This requires a deep understanding of the data and how urban indicators impact geographical areas. However, the tools and mechanisms typically used for this decision-making are generally unsuitable, as they focus on small parts of the data or aggregates. This thesis proposes UrbViz, a visual analytics tool that supports interactive geographical analysis of multi-attribute rankings and causality graphs. The system allows to 1) compare and rank different urban areas regarding index scores and urban attributes, 2) identify hotspots and coldspots on the map for a group of selected indicators, 3) find relationships in geographical areas considering scores of indexes and attributes, 4) analyze and explore causal relations between indicators. The effectiveness of UrbViz was evaluated with real-world urban datasets, including traffic collisions, fire incidents, and neighborhood calls datasets, by two domain experts from a local municipality. The results of the study support the advantages of our tool.

Acknowledgements

I would like to show my gratitude to my supervisor, Dr. Evangelos Milios, for hosting and supporting me in the last part of my research. I am extremely grateful to Dr. Fernando Paulovich for his guidance provided throughout this project and for his warmth in every meeting we had, thank you for the opportunity and for making this journey very exciting. I would also like to extend my gratitude to Dr. Axel Soto, my graduate studies would not have been possible without your help and encouragement over the years.

Many thanks to Halifax Regional Municipality, the partner organization responsible for part of the motivation of this project. Special thanks to Amy Siciliano for this opportunity and all the support she has given to this research. This work was supported by Mitacs through the Mitacs Accelerate program.

None of this would have been possible without the endless support and love of my parents, Walter and Liliana, who always being there for me even though the distance. I want to thank my brother, Matias, for believing in me and feeding my motivation every time it was needed along this journey. I must also thank my amazing friends, from Argentina and Canada, for their immense support and for letting me share this experience with all of them. Finally, I am extremely grateful to James for all the love and patience, and for being my home during my nomad life.

Chapter 1

Introduction

One way to help government entities and municipalities in decision-making is by detecting relationships between indices and urban datasets. This can help to tell a more complete story about the relationship between the entities and the diverse communities it serves. An index is a composite measure to evaluate and understand multidimensional information, aggregating multiple indicators to summarize datasets of different perspectives. It helps identify the strengths and weaknesses of the object under analysis, comparing large data collections through a ranking strategy. In general, indices are deemed the most relevant measure for comparing multivariate data, especially when there is no standard tool for evaluation [37]. Their application includes the measurement of progress in human development, human poverty, and various social and economic indicators [6], [42]. This study uses the Canadian Index of Multiple Deprivation which measures four dimensions of deprivation and marginalization, giving a score to each area. The possibility of finding data patterns between urban datasets and indices allows domain experts to evaluate the impact of concrete variables (some influenced by municipalities' actions) on abstract concepts (e.g., deprivation and marginalization) and to promote hypothesis and exploratory-driven analysis to assist in formulating and answering concrete questions.

Early studies that propose visual analytics approaches for urban data [40], [35], [25], [36] propose a combination of visualizations to analyze the data in detail. However, any of these approaches combine the information from multiple datasets. One way to find relationships between data is by creating rankings to analyze how areas vary in position depending on the indicators they are being ranked. Previous works involve the rank and comparison of different observations regarding multiple attributes, allowing the user to set and change their weights to alter the rank [23], [14], [45], [32], [28]. However, any of these studies make use of urban indices to compare against diverse urban datasets. Other types of approaches, mainly used by

municipalities, involve the spotting of areas with a high concentration of determined urban events [19], [18], [20], [12], [21] with regard to mitigating future effects. However, they focus on one single data sample, not allowing the combination of events from multiple datasets. Analyzing causal relationships, when one event is the result of the occurrence of another event, of different indicators is another way urban data analysts benefit since it can help them to prevent future negative events [48], [52], [49]. However, these studies focus on the causal analysis of attributes from the same dataset, preventing the combination of attributes from multiple datasets.

1.1 Motivation

The motivation for this thesis was to help the Halifax Regional Municipality in finding relationships between their urban datasets and the Canadian Index of Multiple Deprivation. For this, we first derive an analytical workflow and seven user requirements through a careful literature review and close collaboration with domain experts. Based on the requirements, we further developed UrbViz, a visual analytics system to find relationships and patterns among multiple urban data. Motivated by the limitations in the prior studies, UrbViz addresses the gap in the literature as follows. First, we propose two single-column multi-attribute rankings based on an index and numerous urban datasets to rank and compare geographical areas regarding their attributes. Second, we design a dynamic map visualization to spot areas of high/low concentrations of urban indicators based on their past frequency and facilitate identifying relationships between index attributes and urban dataset indicators. Third, we present a geographically located causality network visualization to analyze the causal relation between urban indicators. Finally, a user evaluation with usage scenarios of the system with urban experts demonstrates the effectiveness of our approach.

1.2 Contributions

In brief, our contributions to this thesis are:

- We propose a combination of different techniques to help urban data analysts, including multi-attribute rankings, spatial analyses (hotspots, coldspots, and multivariate choropleth map), and geo-located causal analysis.

- We design and implement a visual analytics system called UrbViz for analyzing multiple urban datasets and indices to identify patterns in the data. UrbViz incorporates multi-attribute rankings, spatial and causal analyses, and a set of effective visualizations for analyzing located geographical areas.
- We evaluate our approach through a user evaluation with usage scenarios with domain experts.

1.3 Thesis Outline

The remainder of this thesis is organized as follows: Chapter 2 presents the literature review of some of the research work previously done about multi-attribute ranking visualizations, urban hotspot analysis, causal relation visualizations and urban visual analytics. Chapter 3 describes the pipeline creation of UrbViz; first, it describes the data used and its pre-processing process, secondly, it presents the requirements set for the system and thirdly, it discusses the methodology of the research describing the system architecture, as well as the data processing pipelines including rankings, spatial and causal analyses, and finally the visual design decisions taken for each visualization in the tool. Chapter 6 presents the user evaluation for the system made with domain experts testing real usage scenarios. Chapter 7 concludes our thesis work by briefly highlighting limitations and future directions.

Chapter 2

Related Work

Research studies that overlap with this work can be divided into four categories: multi-attribute ranking visualizations, urban hotspot analysis, causal relation visualizations, and metropolitan visual analytics. This chapter aims to define the research gap our system intends to fill.

2.1 Multi-attribute ranking visualization

Multi-attribute ranking systems are standard tools used to make data-driven decisions. This type of ranking presents a challenge since it is not trivial how the attributes contribute to a ranking and how changes in one or more attributes influence it. In order to interpret, modify, and compare such rankings, more advanced tools are needed. Researchers have developed different strategies to visualize multi-attribute rankings in the past years. Some studies, for instance, LineUp, Rank As You Go and Relevance Tuner allow the user to create multiple custom rankings by clicking and dragging columns to change the attribute weights, and hence visually show how these changes impact the ranking [23], [14], [45].

An opposite approach, Podium operates in a way in which users do not require to specify the attribute weights, but instead, the user can drag rows and rank data points based on their perception of the relative value of the data. As a consequence, the columns of the attributes show the weight needed to make that ranking possible [47].

Several approaches in the literature are inspired by LineUp and use their ranking as part of the solution to a specific problem. One example is WeightLifter [32], which is an interactive visualization technique for weight-based multi-criteria decision-making that facilitates the exploration of all possible combinations of weights. The approach of SmartAdAp, which is shown in Fig. 2.1, is an interactive visual analytics system to solve billboard placements. The tool deals with two significant challenges: finding

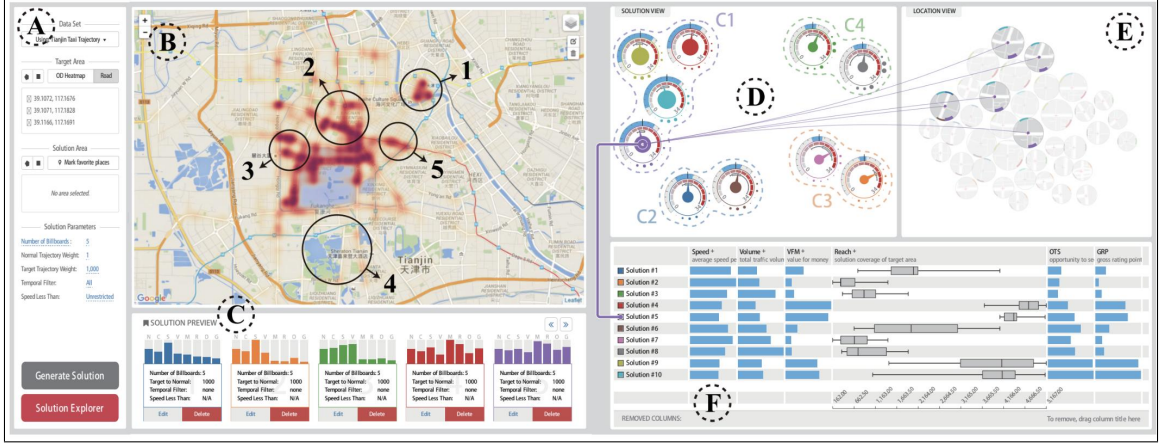


Figure 2.1: Interface of SmartAdP: interactive visual analytics system to solve billboard placements. A framework that integrates a novel visualization-driven data mining model enables advertising planners to effectively and efficiently formulate good candidate solutions.

good solutions in a large solution space and comparing them visually and intuitively [28]. The system WarehouseVis proposes a visualization-driven pipeline to facilitate warehouse location selection. It centralizes relevant information about warehouses and adapts a widely-used methodology used by experts to efficiently rank warehouse candidates [27]. Finally, the study SRVis, shown in Fig. 2.2, proposed a spatial ranking visualization tool that supports efficient spatial multi-criteria decision-making processes. This approach addresses challenges in terms of the presentation of spatial rankings and contexts, their scalability, and the analysis of context-integrated spatial rankings [50].

Figures 2.2 and 2.1 show the integration of multi-attribute rankings into systems that try to solve a specific domain problem. Our approach tries to do the similar by proposing a single-column ranking and a custom one created by combining different urban attributes and interacting together. This technique was inspired by LineUp, and some of its features were used.

2.2 Urban hotspot analysis

Hotspot analysis is a spatial analysis technique frequently used with urban data to describe specific events across different geographical areas. A feature is a hotspot when it has a high value of a specific event, and it is surrounded by other features

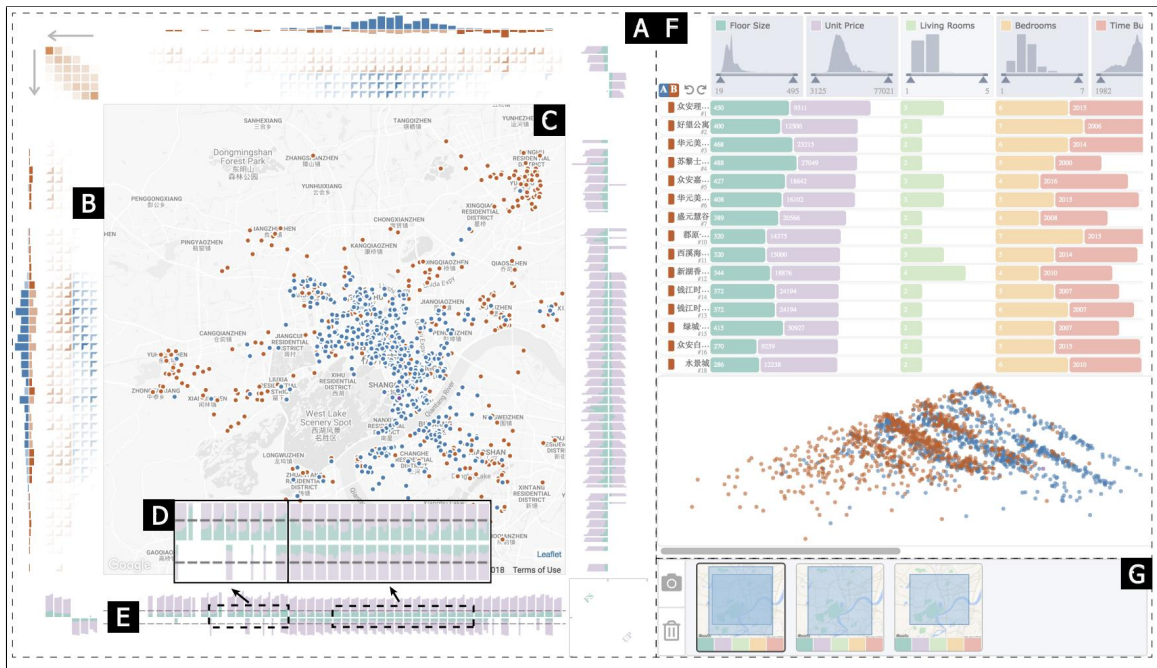


Figure 2.2: Interface of SRVis: a spatial ranking visualization tool that supports efficient spatial multi-criteria decision-making processes.

with a high number of the same event. Crime-related data analysis is one of the most common problems in which hotspot analyses are used. The system CriPAV is a visualization-assisted analytical tool that integrates a set of linked views to uncover relevant information about hotspots. CriPAV enables mechanisms to identify, explore, and analyze crime hotspots in a street-level of detail, enabling users to explore the possible causes of the observed crime patterns [19]. A similar approach is carried out on Mirante, shown in Fig. 2.3, where a crime mapping visualization system that allows spatiotemporal analysis of crime patterns on a street-level scale using street-level heatmaps is proposed [18]. Different from these methods, CrimAnalyzer, shown in Fig. 2.4, proposes a visual analytic tool that allows users to study the behaviour of crimes in specific city regions. The system facilitates the identification of local hotspots and the associated crime patterns while showing how hotspots and corresponding crime patterns change over time [20]. Another example is SHOC, which is a practical tool that facilitates performing domain tasks such as detection, comparison, and evolution of hotspots by providing a straightforward way of making spatial comparisons of areas, using set-like operations on superimposed geometries for visualization [12]. Finally, the system HotSketch allows police officers to rapidly utilize

predictive models for crime in a neighbourhood while away from the precinct. The users can explore crime data in their area through sketch-based interaction methods that enable them to make queries and update a dashboard [21].

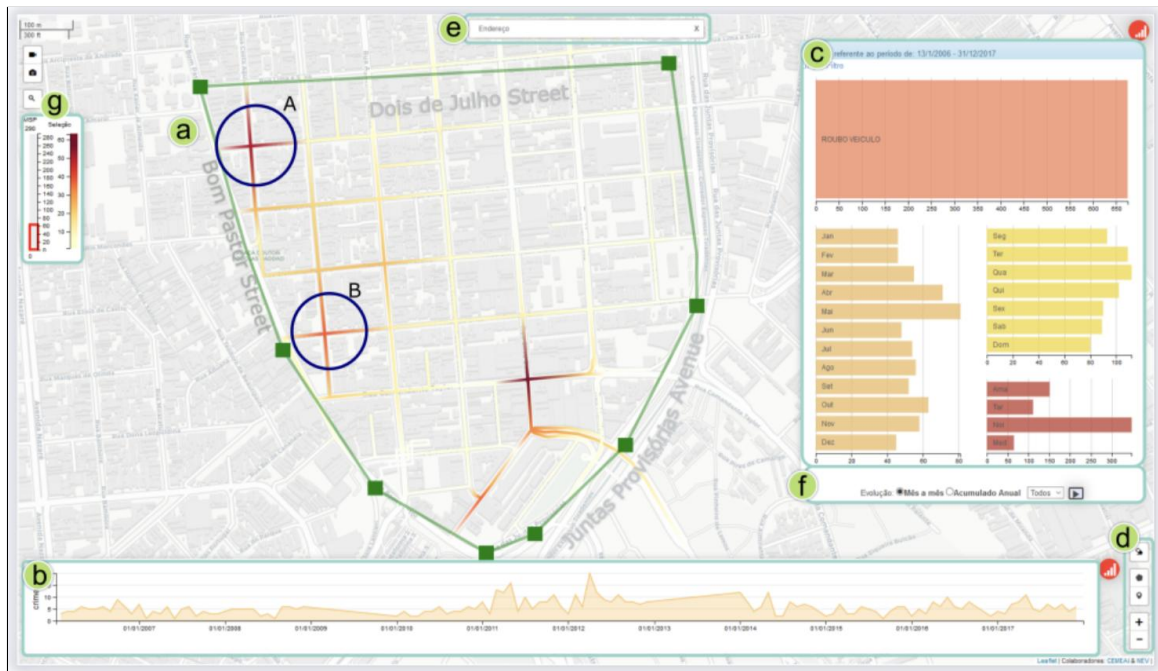


Figure 2.3: Interface of Mirante: crime mapping visualization system that allows spatio-temporal analysis of crime patterns on a street-level scale using street-level heatmaps.

To a lesser extent, some tools use hotspot analysis beyond the crime domain. The study of Garcia et al. presents the development of a GIS system that visualizes traffic hotspots in the road network inside the metropolitan area of Madrid and warns drivers when approaching those hotspots [11]. Another example is the study of Singh et al. which proposes a hotspot analysis method that delineates the areas with the highest vulnerability to fire events [41]. The system can alert urban planners to the need to provide appropriate fire services. The study of Tasseron et al. presents a simple method for plastic hotspot mapping in urban water systems [44]. Similarly, the approach of Cao et al. explores new ways to better understand urban systems emphasizing detection and analysis of urban area hotspots through cell phone traffic data [8].

None of these approaches can combine two or more urban datasets to create a spatial analysis. We can see in Mirante and CrimAnalyzer, shown in Figures 2.3 and

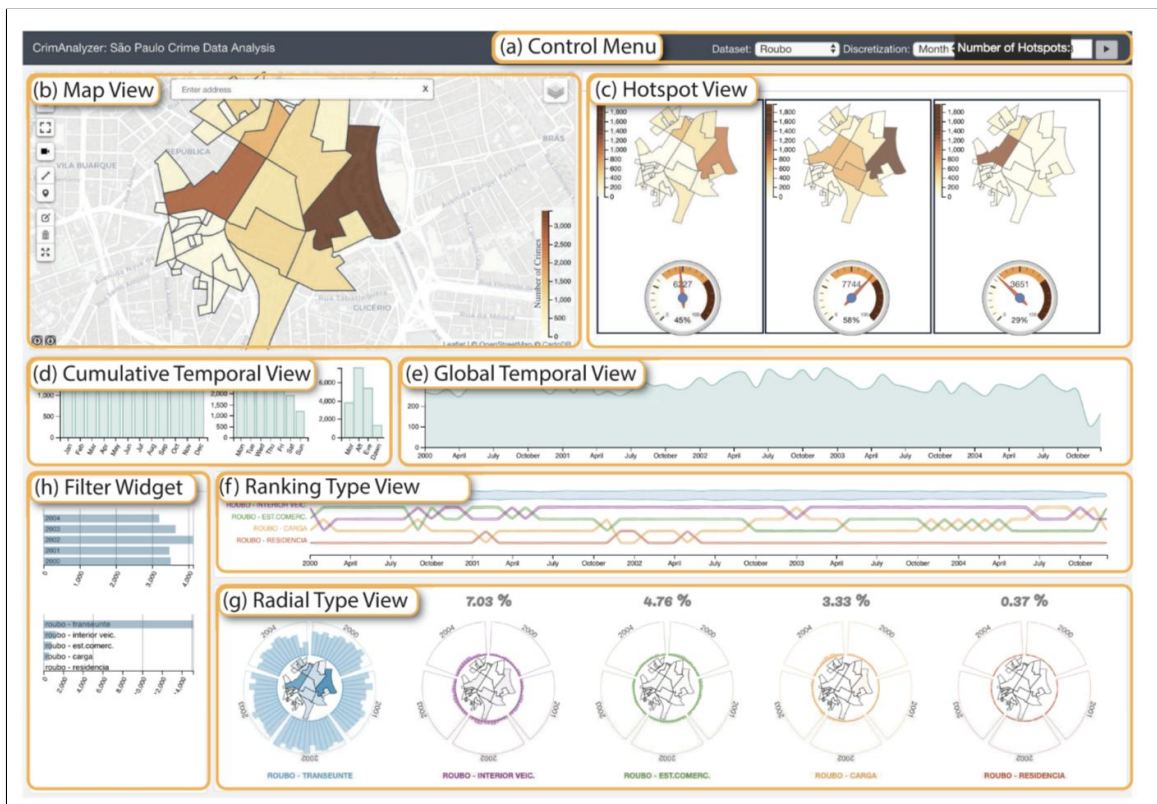


Figure 2.4: CrimAnalyzer System: Visual analytic tool that allows users to study the behaviour of crimes in specific city regions.

2.4 respectively, that their systems focus on one single type of dataset, not offering an analysis of a combination of them. Our approach differs from the previously cited systems in that it combines attributes of different urban datasets chosen by the user to calculate hotspots.

2.3 Causal relation visualizations

A wide variety of methods have been developed to visualize causality relations in data analysis. One is through causality networks, a graph formed by nodes representing variables connected by links denoting causal influence. One example of approaches with causality networks is the system Causality Explorer, shown in Fig. 2.6, which is a visualization tool that allows analysts to explore, validate, and apply causal relations in real-world decision-making scenarios. The tool provides an uncertainty-aware causal graph visualization for presenting a large set of causal relations inferred from high-dimensional data [52]. Similarly, the approach of Visual Causal Analyst shown in Fig. 2.5 is a causal reasoning framework that allows users to apply their expertise, verify and edit causal links, and collaborate with the causal discovery algorithm to identify a valid causal network. It visualizes causal networks as colour-coded 2D graphs with force-directed layouts [48].

Further research by Wang and Mueller resulted in the creation of Causal Structure Investigator, an interface that includes visual analytics techniques for making visual causality analysis more practical for real-world applications [49]. Instead of causal networks, other approaches use event sequence visualizations to explain events' occurrence. The study by Xie et al. proposed an interactive visual analytics approach that combines Sankey diagrams and timeline-based visualizations to present the causality of event sequences [53]. Similarly, Compass is a visual analytics approach for in-depth analyses of dynamic causality in urban time series. This tool uses dynamic graph visualization to reveal the temporal variation of the detected causal relations and enables time-oriented drill-down analyses of causal graphs [13].

Other methods to visualize causal relations exist. For example, the study by Yafeng et al. explore linking textual media data with curated secondary textual data sources [29]. Relationships between these events and media trends can be analyzed using causality modeling, where model results are interactively displayed. Due to

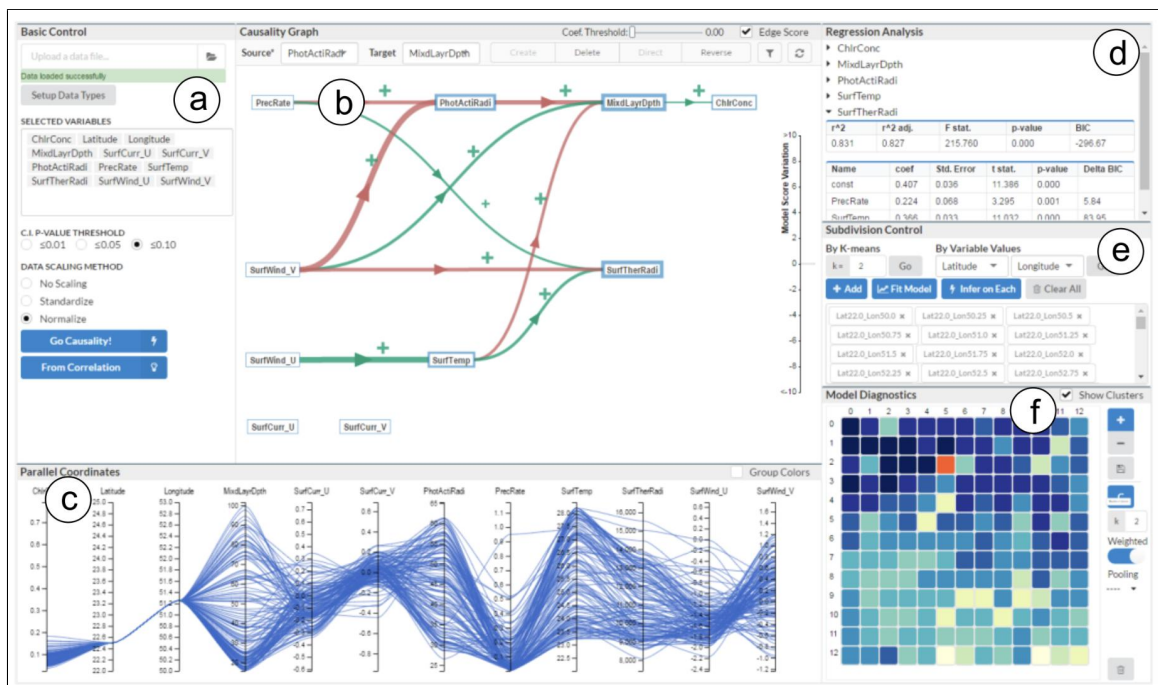


Figure 2.5: Visual Causality Analyst framework: reasoning framework that allows users to apply their expertise, verify and edit causal links, and collaborate with the causal discovery algorithm to identify a valid causal network. Its interface consists of both an interactive 2D graph view and a numerical presentation of salient statistical parameters, such as regression coefficients, p-values, and others.

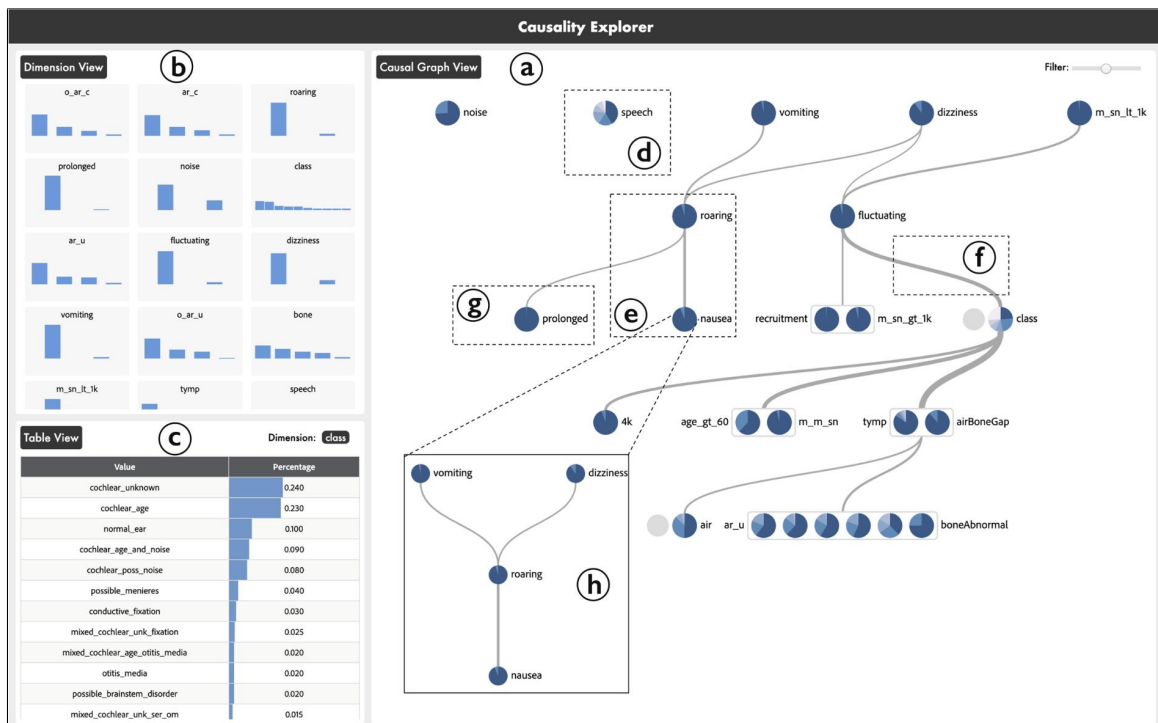


Figure 2.6: Interface of Causality Explorer: A visualization tool that allows analysts to explore, validate, and apply causal relations in real-world decision-making scenarios. The tool provides an uncertainty-aware causal graph visualization for presenting a large set of causal relations inferred from high-dimensional data.

the wide adoption of node-link diagrams for understanding mediated causation along causal paths [46], our tool also uses this visual metaphor considering that dynamic causality in urban space integrating different topics has not been studied. In Fig. 2.6 and Fig. 2.5 show the approaches of Causality Explorer and Visual Causality Analyst respectively. We can observe both of them offer a causality network for different topics but the option of geographically located causality is not offered. Our approach creates causal networks of causal relations for different geographical locations between different combinations of urban dataset attributes.

2.4 Visual urban analytics

Visual analytics approaches have been proven effective in leveraging various urban data to improve our understanding of cities. In this section, we restrict ourselves to approaches that handle spatiotemporal datasets to solve problems for urban-data experts. The study by Seebacher et al. proposes an interactive visual analytics method that consolidates automatic predictive analysis and glyph-based overview visualizations. This tool aims to support domain experts, such as city planners or meteorologists, in detecting and investigating urban heat islands [38]. Another approach with the same goal is the interactive visualization system StreetExplorer, which is used for the exploration of global and local properties of urban street networks to help urban morphology experts identify patterns of historical development and modern planning approaches [40]. Urban Chronicles is an open-source web-based visual analytics system that enables interactive exploration of changes in land use patterns[35]. TrajGraph is another visual analytics system for studying urban mobility patterns by integrating graph modeling and visual analysis with taxi trajectory data [25]. The graph is created to store and display real traffic information recorded by taxi trajectories over city streets. The final approach, VisMiFlow, proposes a visual analytics system to support and improve city planning decisions based on city migration movements [36].

Most of these tools focus on a single urban dataset to analyze such data visually and interactively. One exception is the system Urban Pulse [30], which proposes a data-driven approach to understanding the city in the context of multiple urban data sets to determine a city’s level of activity. VAUD [9] is a visual analytics system

that supports visual queries and reasoning across multiple semantically linked urban data sets. Our work can use multiple datasets to make different computations to find patterns between them.

2.5 Summary

In summary, UrbViz proposes a novel combination of visualizations to solve a domain problem. This approach uses multiple datasets to make different computations to find patterns between them. It allows the users to perform individual analyses such as hotspots, coldspots, and causal relations, and also the option of combining different datasets in rankings and multivariate choropleth maps. We have not found many other approaches that offer this combination of techniques. Also, none of the works previously cited meet the whole set of domain tasks and requirements that arise from our collaboration with domain experts.

Chapter 3

UrbViz

This chapter describes the creation process of UrbViz. First, it is described the datasets used in the system and its preprocessing pipeline. Second, it is presented the requirements for the system. Finally, it is explained the system architecture, as well as the implementation of all features and visual design decisions.

3.1 Datasets

The datasets used in this thesis correspond to the data collected by Halifax Regional Municipality (HRM) during the last 10 years regarding different urban topics. The Canadian Index of Multiple Deprivation(CIMD), openly available was also added as part of the system data [43]. Our data comprises:

- **Traffic collisions.** Point representation of traffic collisions that have occurred within the road right-of-way. Collisions are mapped according to the Motor Vehicle Collision Report. It includes all the traffic accidents that occurred in the municipal region during 2018 - 2022.
- **Fire incidents.** Includes all the fire incidents that occurred in the municipal area during 2006 - 2021.
- **Neighbour calls.** Collection of phone calls made by the citizens to the Municipal Services and Information Center, a free municipal government service that provides information from anywhere in the region on municipal information and services. The entries are calls made from 2017 to 2021.
- **Parks and Public buildings.** Pointwise location of parks inside the municipality district and public buildings such as hospitals, universities, supermarkets, and so on.

- **Canadian Index of Multiple Deprivation.** Area-based index of the 2016 Canadian Census of Population microdata to measure four key deprivation scores at the dissemination area (DA) level: residential instability, economic dependency, situational vulnerability, and ethnocultural composition. Each score is computed by an unknown formula using different indicators.

All dataset entries contain the dissemination area code where the event happened. A dissemination area (DA) is a small, relatively stable geographic unit composed of one or more adjacent dissemination blocks. A dissemination block (DB) is an area bounded on all sides by roads and boundaries of standard geographic areas. DAs are uniform in terms of population size, which is targeted from 400 to 700 persons to avoid data suppression. This means that the regions have different sizes. For instance, DAs in downtown are smaller than the ones in the suburbs since there are more densely populated. Instead of using DAs, we segment our data into polygons (hexagons in this case) covering the geographic area under analysis. The concept of *area* used in this study will make reference to a hexagon polygon. All hexagons have the same size, making it possible to compute the spatial analyses since this is the proper scale for the type of application we support (later discussed). To generate the hexagon map, we preprocessed the datasets by adding a hexagon ID according to the geolocation of the record. In the case of the Canadian Index of Multiple Deprivation, when we place the hexagon layer on top of the DA layer, some hexagons cover one or more DAs. This means that the index score, for instance, for *Residential Instability* of a hexagon is computed considering all scores of *Residential Instability* of the overlapping DAs. To solve this, we decided that the index scores for each hexagon are the weighted average of the different DA scores covered based on the percentage of surface area the hexagon is covering.

3.2 Requirements

Urban data analysts working on municipalities spend considerable time finding patterns and possible relationships in data to find insights and the reasons underpinning urban behaviors. We have been collaborating closely with a municipality's experts to understand their needs and provide them with a tool for data analysis. We had

multiple interviews with urban data analysts to identify the problems they face during data analysis and brainstorm possible solutions. During this process, we analyzed their needs and elicited requirements for a software tool to help them in their tasks. After identifying the experts' requirements, we collected relevant research articles and developed a prototype system. Finally, we verified and refined the user requirements together with the prototype system through interviews with the experts. After the structured interviews with the domain experts, we compiled a list of requirements based on their analytical needs:

- R.1 **Encode multi-attribute ranking:** Domain experts typically work with multiple indicators, which must be weighted and combined to produce a single ranked list to prioritize citizen demands. Therefore, an analytical tool must allow the user to create a ranking from a chosen dataset and combine multiple indicators to produce a single all-encompassing ranking.
- R.2 **Correlation between index scores and ranking attributes:** Domain experts require to rank and compare areas of the region based on the score index and attribute(s) from the urban datasets they have. The system must support comparing data items in different ranking positions to find correlations between the custom-made ranking indicators and the Canadian Index of Multiple Deprivation scores.
- R.3 **Relationship between index and data:** Domain experts waste a considerable amount of time in identifying where in the map there are relationships between an index and the different urban datasets. The experts require an easy and quick way to identify spatial patterns or relationships between the index and datasets on the map.
- R.4 **Highlights in the map:** Domain experts require a technique that allows them to spot easily on the map the places with the highest/lowest correlation between an index score and urban attribute(s). The system should allow the user to explore the polygon areas of interest from the custom-made ranking on a map. Depending on the data and hypothesis under analysis, this exploration should be possible at different spatial granularity levels.

R.5 Hot and cold spots: Domain experts need techniques that help them to identify spots that contain a high frequency of certain events to mitigate possible future negative effects. The system should identify the geographic areas where it is likely and not likely that events will repeat based on frequency in historical data.

R.6 Causal relation among indicators: Domain experts need to identify which events trigger future events to mitigate them. The system should allow users to analyze causal graphs for different areas in detail, including the directions and paths of the causal relations among all attributes.

R.7 Location of entities: Domain experts typically use city landmarks for both reference and hypothesis generation. Therefore, the system should support the pointwise location of parks and important buildings (e.g., schools, hospitals, etc.) and the possibility of exploring any correlation with the spatial layer of hotspots, coldspots, and multivariate choropleth map.

3.3 System

In this section, we first describe the architecture of our tool. Then, we detail the spatial and causal analyses that are performed in the back-end. Finally, we describe the visual encodings and different views used in our interface.

3.3.1 System architecture

The system is a web-based application constituted by a data storage, its back-end and front-end. The back-end is coded in Python and it handles the pre-processing of data and the calculation of the spatial and causality tests. The front-end is developed in Svelte.js [24] with the support of libraries such as D3.js [7] and Leaflet.js [1]. The front-end helps users find patterns in the data by using hotspots, coldspots, and multivariate choropleth maps. It also supports the analysis of causal relations with the causal graph visualization.

3.3.2 Data processing pipelines

This section introduces the proposed methods to compute the main functionalities of UrbViz.

Ranking and indices

Based on our requirements, experts need a flexible ranking tool to create rankings from scratch using their available datasets (R.1). UrbViz shows two rankings: Ranking 0 and Ranking 1 in the Ranking View as it is shown in Fig. 3.1. Both are *single-column multi-attribute rankings*, i.e., they are composed of one single column created by different sources:

- *Ranking 0*: A single column ranking of the hexagon areas made from an index chosen by the user. As the scores of the index are calculated by an unknown formula, we cannot join or combine attributes, so only one score is shown.
- *Ranking 1*: A single weighted sum column ranking of the hexagon areas that combines multiple attributes chosen by the user.

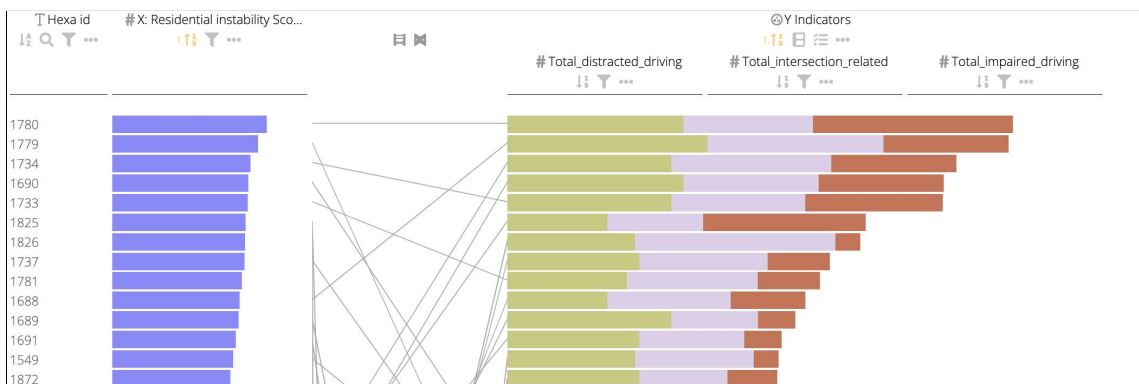


Figure 3.1: Rakings 0 and 1 in the Ranking View.

They also require visualizing and connecting the ranked position of the same data item in the second ranking (R.2). The ranking technique that we used was inspired by LineUp [23], as we used some of its features. In Ranking 1 the user assigns a weight w to each indicator chosen by the user according to her assessment of its importance in the ranking. Initially, the weights are all equal. As the user changes

the weights, Ranking 1 gets updated dynamically. The score for each data item is computed as a weighted sum of its individual attribute scores. The such score is used to define the position of the ranked items. Figure 3.2 A' shows the user selected three desired indicators for Ranking 1, and the weighted sum is computed among them. Since the attributes can be of different types, i.e., continuous and categorical, the user needs to normalize the attribute vectors by mapping them onto a common scale. For numerical attributes, the user provides a mapping function stating each attribute's minimum and maximum values. For categorical attributes, their order will determine the ranking of the observations.

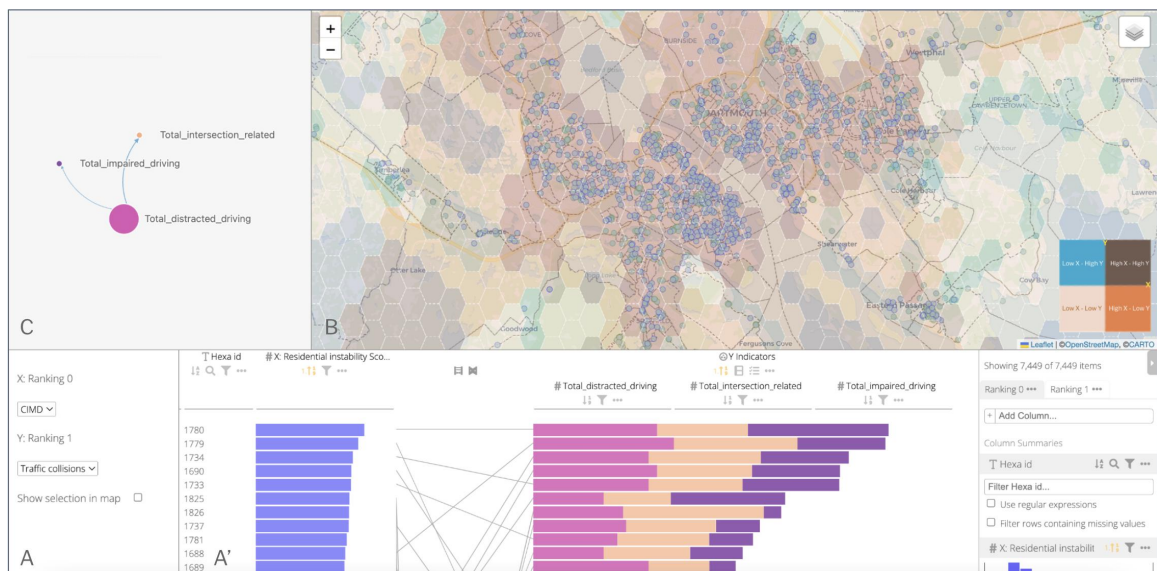


Figure 3.2: UrbViz contains three main views: (A) Ranking View, (B) Map View, and (C) Causality View. The Ranking View has 2 sub views: (A') is the menu ranking where the users select the datasets they want to work with and (A'') presents two single-column rankings, one from the CIMD and the second from an urban dataset. The Map View shows a map with five different layers: hotspots, coldspots, relationship map, and pointwise locations of buildings and parks. The Causality View displays a causal network composed of nodes representing the indicators from Ranking 1, which are connected by links that denote causal influence.

Spatial analyses

Spatial analyses allow us to solve complex location-oriented problems, explore and understand data from a geographic perspective (R.4), determine relationships among

different areas (R.3), find patterns, assess trends, and make decisions based on geo-located data (R.5). In this domain, visualization of spatial analyses is a key part of our tool as it provides the experts with a convenient way to spot patterns. This section explains the computations for each spatial analysis included in the visualization tool.

Hotspots and Coldspots As illustrated in Figure 3.2 B, hotspot and coldspot identification is a primary component of UrbViz. These spots are identified by local measures of each observation and its surroundings. A hotspot is shown for an area of the map based on a geo-located event or attribute e , when e is significantly higher than the mean and it is also surrounded by other high values of e [33]. There may be areas on the map with high values of e , but if other high-value areas do not surround them, it is not regarded as a hotspot. On the contrary, coldspots are areas with low values of e compared to the overall mean that is also surrounded by other areas with low values of e . The hotspots and coldspots are calculated on the hexagon polygons, their visualizations are shown in Fig. 3.3.

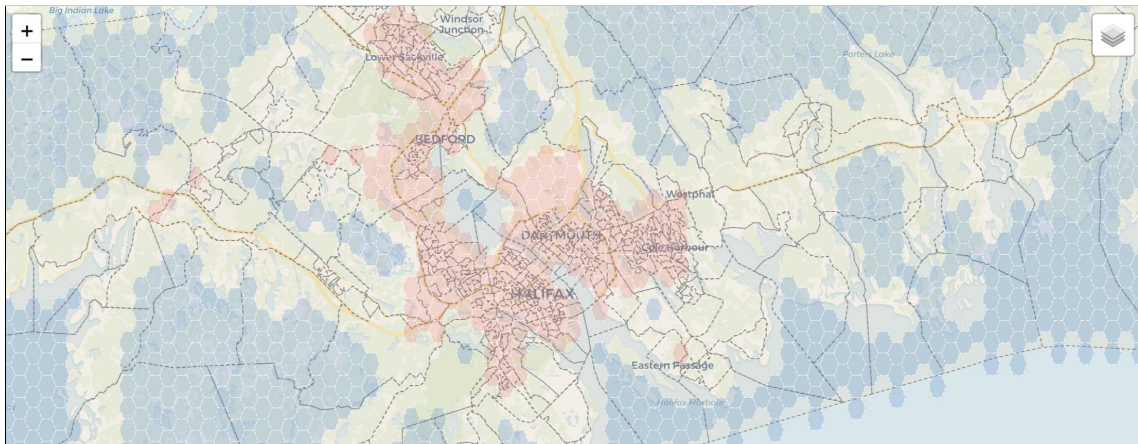


Figure 3.3: Hotspots (Red hexagons) and Coldspots (Blue hexagons) shown in the Map View.

The identification of these spots includes the analysis of a variable behavior in a particular area, and because nearby areas tend to have similar behaviours, we also have to consider them in the analysis. This is called spatial autocorrelation. Spatial autocorrelation has to do with the degree to which the similarity in values between observations in a dataset is related to the similarity in locations of such observations. It relates the value of the variable of interest in a given location with values of the same

variable in other locations. This type of autocorrelation can be calculated depending on the scale we are considering, i.e., global or local. Global spatial autocorrelation considers the overall trend that the location of values follows. Local measures of spatial autocorrelation focus on the relationships between each observation and its surroundings rather than providing a single summary of these relationships across the map [39]. In order to calculate these correlations, we compute a spatial weight matrix to represent the spatial relationships. Spatial weights connect objects in a geographic table to one another using the spatial relationships between them. Adding all of them in a matrix lets us embed our observations in space together. The weights express the neighbor structure between the observations as a $n \times n$ matrix W in which elements w_{ij} of the matrix are the spatial weight. Spatial weights w_{ij} are non-zero when i and j are neighbours, and zero otherwise.

In calculating hot and cold spots, testing the presence of clustering in the data by using global spatial autocorrelation techniques is the first step. In our approach, we decided to use Moran's I [31], which given a set of hexagon areas and an associated attribute, evaluates whether the pattern expressed is clustered, dispersed, or random [34]. Moran's I can be written as:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2}, \quad (3.1)$$

where n is the number of observations, z_i is the standardized value of the variable of interest at location i , and w_{ij} is the cell corresponding to the i -th row and j -th column of a spatial weight matrix W . Moran's I returns a value ranging from -1 to 1. Perfect dispersion is at -1, complete random arrangement at 0, and a north/south division at +1, indicating perfect correlation. For each attribute we have in the datasets, we calculated Moran's I score and a pseudo-p-value called *p_sim*. *P_sim* is calculated in order to answer the questions of *how likely is the pattern we observe in the map and Moran's I captures in its value to be generated by an entirely random process?*. It is calculated as an empirical p-value representing the proportion of realizations in the simulation under spatial randomness that are more extreme than the observed value. A small p-value, below 0.05, associated with Moran's I of a map allows rejecting the hypothesis that the distribution of values of an attribute in the map is random.

After these calculations, we captured each indicator's degree of geographical clustering. To calculate where the specific clusters are on the map, we calculate the

Local Indicator of Spatial Association (LISA). LISA has two important characteristics. First, it provides a statistic for each location with an assessment of significance. Secondly, it establishes a proportional relationship between the sum of the local statistics and a corresponding global statistic [3]. In our approach, we decided to work with the LISA statistic Local Moran's I, which is used to identify local clusters and local spatial outliers. The core idea of a Local Moran's I_i is to identify cases in which the value of an observation and the average of its surroundings is either more similar or dissimilar than expected from pure chance. The mechanism to do this is similar to the one in the Global Moran's I but applied in this case to each observation. The formal representation of the statistic can be written as follows:

$$I_i = \frac{z_i}{m_2} \sum_j w_{ij} z_j; m_2 = \frac{\sum_i z_i^2}{n} \quad (3.2)$$

where m_2 is the second moment (variance) of the distribution of values in the data, $z_i = y_i - \bar{y}$, w_{ij} , is the spatial weight for the pair of observations i and j , and n is the number of observations [39]. After calculating the Local Moran's I value for each observation, we can use a Moran Scatterplot to analyze and explore the results. A Moran Scatterplot consists of a plot with the spatially lagged variable on the y-axis and the original variable on the x-axis [5]. A spatially lagged variable is a weighted sum or a weighted average of the neighboring values for that variable [2]. The x-axis represents the values of an indicator at location i , while the y-axis represents the values of the same indicator in the neighbourhood of location i . Therefore, values in the top-right of the scatterplot represent locations in which the attribute at i and its neighbours are well above the mean, i.e, a High-High relationship, indicating positive spatial autocorrelation. Points in the lower corner of the plot indicate locations in which the attribute at i and its neighbours are below the mean (Low-Low); this also represents locations of positive spatial autocorrelation because nearby locations are similar. The two other quadrants are spatial outliers (High-Low / Low-High) because they are points whose locations are not typical of the patterns of the rest of the points in the dataset. We calculated the local Moran's I for all indicators we have in each dataset and then plotted the results using the Moran Scatterplot to show the relationship between each location and its mean.

Because this analysis identifies statistically *significant* hotspots and coldspots, we

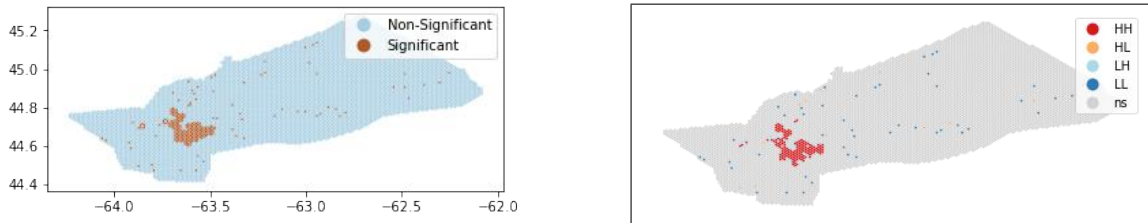


Figure 3.4: Fig. (Left) shows the significance map for the indicators chosen. Fig. (Right) shows the cluster map for the indicators chosen.

need to assess the significance value. For this calculation, we use the conditional permutation method, which calculates a reference distribution for the statistic under the null hypothesis of spatial randomness by randomly permuting the observed values over different locations. The statistic is computed for each randomly reshuffled data set, yielding a reference distribution. The result is a pseudo p -value for each location, which can then be used to assess significance. When an indication of significance is combined with the location of each observation in the Moran Scatterplot, a powerful interpretation becomes possible. The combined information allows for classifying significant locations as High-High and Low-Low spatial clusters, and High-Low and Low-High spatial outliers. The significance map in Fig. 3.4 (Left), which is a representation of positions of significant transform coefficients, shows the locations with significant local statistics for the chosen indicators, starting with $p < 0.05$, and shows all the meaningful categories for the given number of permutations. The cluster map shown in Fig. 3.4 (Right) of the same indicators augments the significant locations with an indication of the type of spatial association based on the location of the value and its spatial lag in the Moran scatter plot. Observing the significance and cluster map we can conclude that positive forms of local spatial autocorrelation are of two types. First, HH observations, which we can term *hotspots*, represent areas where values at the site and its surroundings are larger than average. Second, LL observations, significant clusters of low values surrounded by low values, are sometimes referred to as *coldspots*. Negative forms of local spatial autocorrelation also include two cases. When the focal observation displays low values but its surroundings have high values (LH), we call them *doughnuts*. Conversely, areas with high values but neighbored by others with low values (HL) can be referred to as *diamonds in the rough*[4].

Relationship Map. Multivariate choropleth maps can help experts find patterns between the indicators from the urban datasets and the scores from the Canadian Index of Multiple Deprivation (R.3). For this reason, we decided to include a multivariate choropleth map since they allow the visual comparison of multiple attributes simultaneously. In essence, a relationship map combines two data attributes to show where they converge (have similar values) and where they diverge (have opposite values) [16].

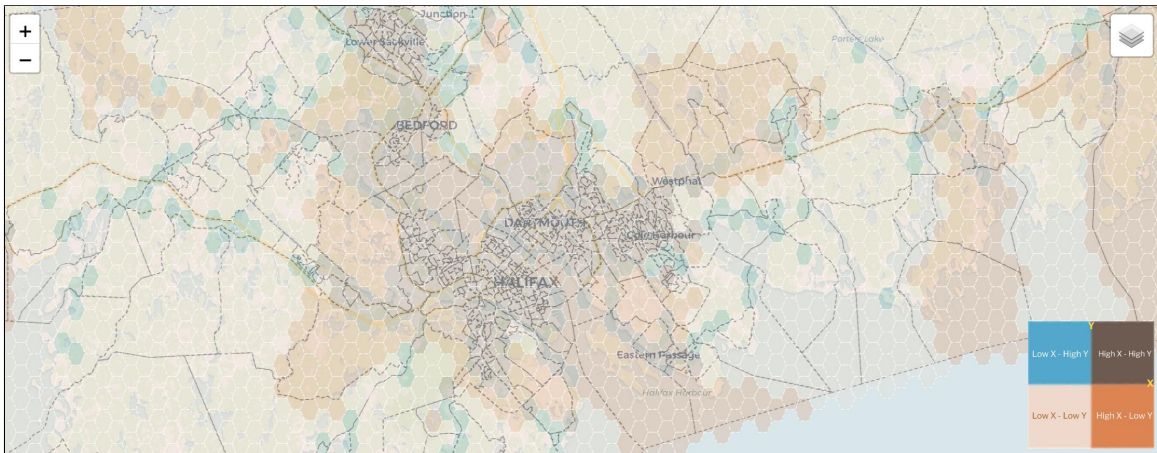


Figure 3.5: Relationship Map layer shown in the Map View.

For instance, if the user selects the column “Residential Instability Score” for Ranking 0, and “Total intersection related and Total distracted driving” for Ranking 1, then the tool creates the sets $X = \text{Residential Instability Scores}$ and $Y = \text{Total intersection related} + \text{Total distracted driving}$. UrbViz then standardizes the values of X and the sum of the columns included in Y . Once the data is standardized, the tool displays the plot $RM(X, Y)$ in a fourth quadrants for all the hexagons we have on our map. $RM_i(X, Y)$ can be described as a point representing the standardized X and Y values for the hexagon i . Once the tool calculates all $RM(X, Y)$, the results are plotted on a scatterplot to analyze the data and continue with the process.

Emulating a similar approach to the Moran Scatterplot, we divided the plot into four (4) quadrants. The top-right quadrant represents the hexagons with a high score of X and a high number of Y . These are the areas called High-High (HH). The bottom-left shows the Low-Low (LL) areas, hexagons with low scores of X and low numbers of Y . The scatterplot also displays spatial outliers, areas with a high score

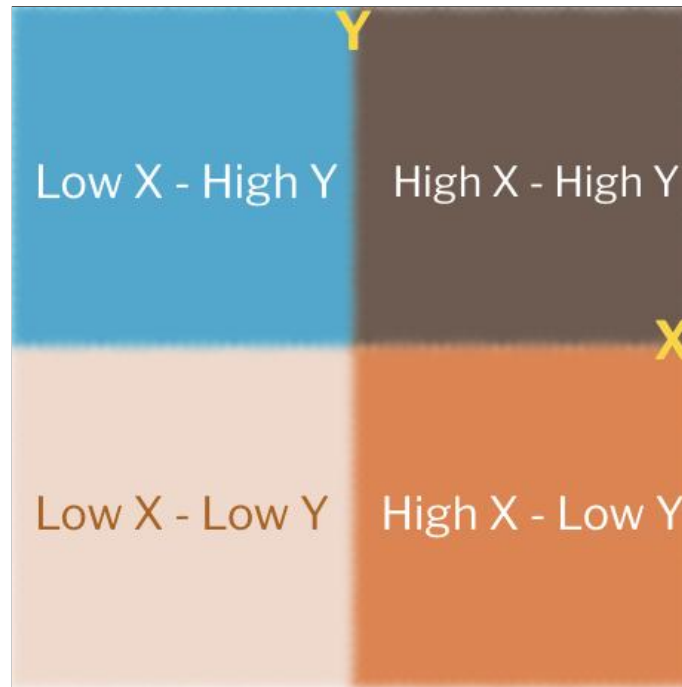


Figure 3.6: Legend (Color scale) of the Relationship Map.

of X but a low number of Y are plotted in the top-left quadrant, the High-Low (HL) areas. The Low-High (LH) areas are shown in the bottom-right quadrant. We decided to work with four (4) quadrants, assigning to each of them a specific colour. Fig. 3.5 shows the visualization for the relationship map and Fig. 3.6 shows the grid colour used. The colors we decided to use in this approach are:

- Brown for HH: Areas with high score from a column of the CIMD and high number of urban attributes chosen.
 - Blue for HL: Areas with high score from a column of the CIMD and low number of urban attributes chosen.
 - Orange for LH: Areas with low score from a column of the CIMD and high number of urban attributes chosen.
 - Beige for LL: Areas with low score from a column of the CIMD and low number of urban attributes chosen.
- . All the points (hexagons) receive the colour assigned to the quadrant they belong.

Finally, we assign the quadrant colour to each hexagon in the hexagon layer. Now, the user can analyze the map and have a quick and easy overview of the relationships between the index scores and indicators of the dataset chosen.

Location of entities Domain experts from Halifax Regional Municipality expressed the importance of overlapping the different spatial analyses with the location of important parks and buildings from the city to find new correlations with these entities (R.7). For this reason, we include two layers with the pointwise location of parks and buildings shown in Fig. 3.7, respectively, and the option of overlapping them with the rest of the layers.

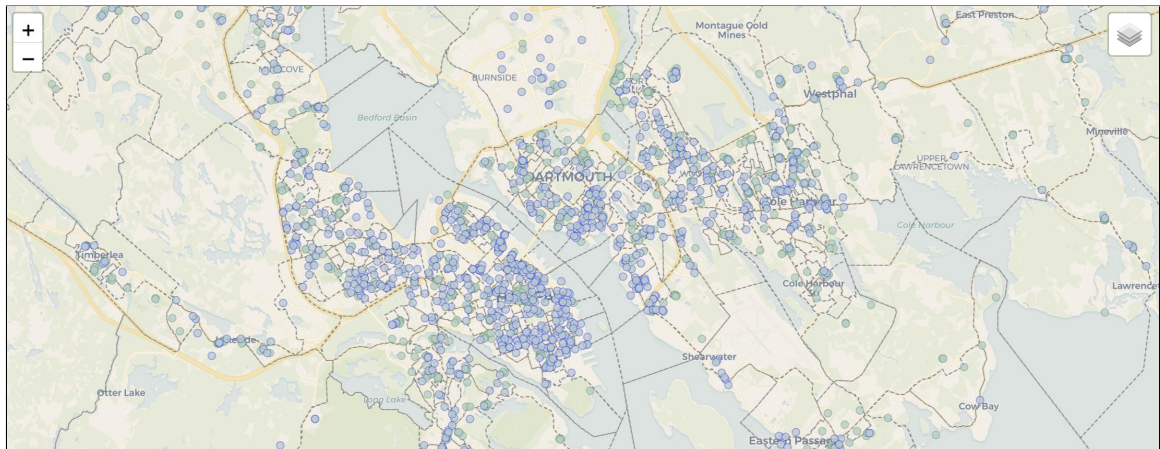


Figure 3.7: Pointwise location of parks and buildings in the Map View.

Causality tests

The possibility of predicting future events becomes vital in the context of municipal entities since it can help urban data analysts make better decisions, e.g., for financial planning or for resource management.

UrbViz contains diverse techniques and visualizations to analyze and identify patterns among data, but since “correlation does not imply causation”, we decided to incorporate the notion of causality through a causal graph to analyze the causal influence between the indicators of the same ranking. Causality refers to the relationship between events where one set of events (the effects) is a direct consequence of another set of events (the causes). Causal graphs have the potential to transform local

government events into a visualization that illustrates likely outcomes and identifies where needs must be met (R.5). A causal graph $G = (V, E)$ is a directed acyclic graph that consists of a set of vertices V denoting the indicators chosen by the user in the custom-made ranking and a set of directed edges E denoting the causal relations between two variables. This type of graph is shown in Fig. 3.8. The basic graph patterns of the causal relations among any three observed variables related to each other are: (1) a common cause influencing multiple variables, (2) a chain of causal influences, or (3) a common effect caused by multiple variables [48].

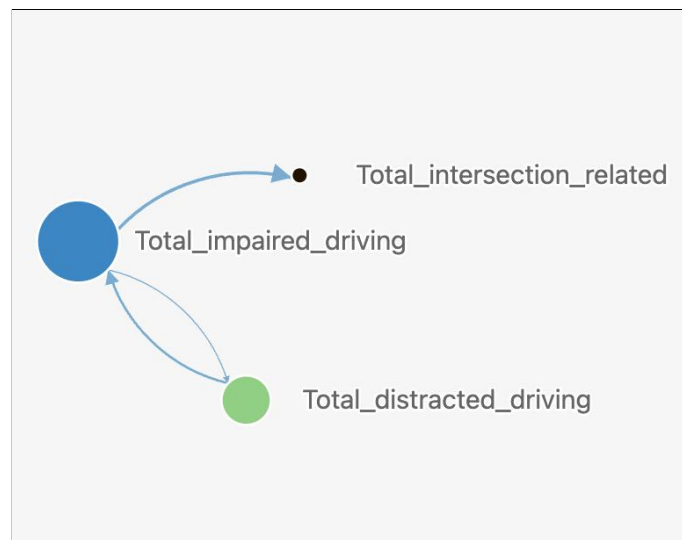


Figure 3.8: A causal graph shown in the Causality View.

The causal relations were calculated by running the Granger causality test [22] among the selected indicators in Ranking 1. The Granger causality test is a statistical hypothesis test for determining whether one time series is a factor and offers useful information in forecasting another time series. For $V = \{v_1, v_2, \dots, v_n\}$ the test was first run selecting v_1 and using the pairs (v_1, v_i) $v_i \in V, i \neq 1$. This step was repeated until all the different pairs of combinations were calculated. Before running the Granger causality tests, a series of steps have to be followed in order to ensure correct causal analysis. First, we preprocess our datasets so that we keep: (1) the date on which events were recorded; (2) the Id of the hexagon area where events took place; and (3) the number of events that happened in the hexagon area on that date.

A prerequisite for performing the Granger Causality test is that data need to be stationary. This means that specific attributes of the data do not change over time.

The augmented Dickey-Fuller (ADF) test is a statistical test that works based on the unit root test to check whether a time series is stationary [15].

After applying the ADF test, the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) is evaluated to determine whether a time series is stationary around a mean or linear trend, or whether it is non-stationary due to a unit root [51].

After performing the ADF and KPSS tests, we can conclude whether the data is stationary. Non-stationary or trending data can often be transformed into stationary by the first-differencing method [26].

The next step is to calculate the vector autoregression (VAR) model [10]. VAR is a statistical model that captures the relationship between multiple quantities as they change over time. VAR models generalize the single-variable (univariate) autoregressive model by allowing multivariate time series. The result of this step is a correlation matrix for residuals for the different time series chosen. We then run the Durbin-Watson test, a measure of autocorrelation in residuals from regression analysis. A Durbin-Watson result of 2.0 indicates no autocorrelation detected in the sample, values from 0 to less than 2 points to positive autocorrelation, and values from 2 to 4 mean negative autocorrelation [17].

After these calculations, we can finally run the Granger Causality Test. For this, we

	Total_aggressive_driving	Total_distracted_driving
Total_aggressive_driving	1.0	0.1449
Total_distracted_driving	0.0025	1.0

Table 3.1: Causal matrix for *total_distracted_driving* and *total_aggressive_driving* from traffic collisions dataset

created a Granger causation matrix $n \times n$ where n is the number of indicators chosen from the custom-made ranking. For each cell in the matrix, we run the Granger causality tests. The resultant matrix of this example is shown in Table 3.1. The row is the response, and the columns are the predictors. If a given p-value is less than the significance level (0.05). For example, taking the value 0.0025 in Table 3.1, we can reject the null hypothesis and conclude that *Total distracted driving* does not Granger cause *Total aggressive driving*. In contrast, the 0.1449 in (row 1, column 2) refers to *Total aggressive driving* Granger causes *Total distracted driving*.

3.3.3 Visual Design

The following section explains the design decisions taken to create the different views in the system.

Overview

The system comprises three views: Ranking view, Map view, and Causality view. The views interact together in most cases, any change in one of them affects the other two. The user selects from the Ranking view the datasets which are desired to work with. Once Ranking 1 and Ranking 0 are made, the user can inspect the Map view to observe the hotspots and coldspots of the indicators from the columns chosen in Ranking 1, and also analyze the relationship map where each axis is represented by the columns of each ranking. Moreover, the user can inspect in the Causality view the causal graph for the entire region in order to see any causal relation between the indicators of Ranking 1. In case a particular area of the municipal area needs to be analyzed in detail the user can select the areas of interest in the Ranking view and show them on the map; a causal graph of the selected area is created in the Causality view, as it is shown in Fig. 3.9.

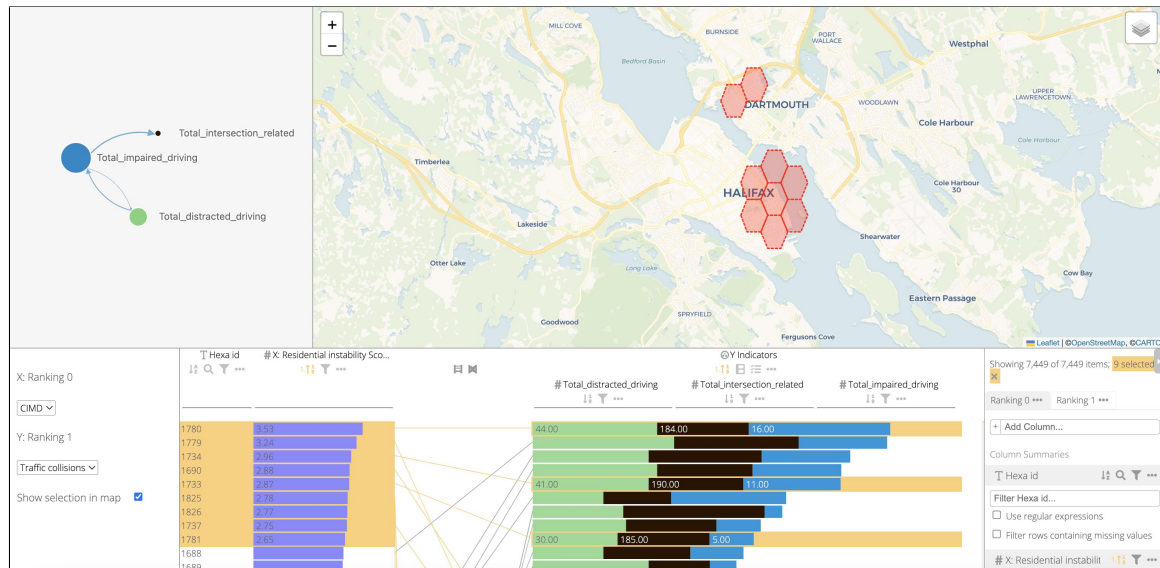


Figure 3.9: First 9 rows of Ranking 0 are highlighted in Ranking 1, and the visualizations of the map and causal graph are updated with the information of these areas.

Ranking view

In this view, the user must select the two datasets needed to create Ranking 0 and 1 in the ranking menu shown in Fig. 3.2 A. The first must be an index, and the second should be any urban dataset with indicators. These two datasets must share the same type of geo-localization for each area they contain, for instance, geographic coordinates or postal code. Once the data is selected, the system creates two single-column rankings. On the left of the view, Ranking 0 shows a column of one of the scores, and on the right Ranking 1 makes a weighted sum of all columns selected by the user (R.1), the rankings are shown in Fig. 3.2 A'. Between the rankings there are parallel coordinates indicating the positions of the data items in both of them (R.2). The attributes have assigned weights that initially are equal for all of them. Altering weights can be realized either by changing the width of a column using drag-and-drop or by double-clicking on the percentages shown above the histograms to specify the exact distribution of weights. While the combined scores are represented as stacked bars, the weight of an attribute is directly reflected by the width of its column header and histogram [23]. It is possible to replace the stacked bars shown in the rankings for each data item with another type of visualization, such as proportional symbols, ticks, strings (of the scores), dots, etc. The first column lists the areas after the two rankings are sorted in descending order, this is visualized in Fig. 3.2 A'. The rest of the columns display the indicators values, encoded by the bars' length. The width of a column represents the weight users assign to the attribute. The colour scheme used for the bars is consistent with the nodes in the causal graph in the Causality view. Users can click on the header of a column to rank the solutions by the associated indicator. In the case of Ranking 1, the columns can also be sorted by any individual column included in the weighted sum. Selection and highlighting of multiple data items from one of the rankings will automatically highlight the same items on the other one, as it is shown in the Ranking View in Fig. 3.9.

Map view

The Map view in Fig. 3.2 B visualizes an OpenStreetMap centred in the municipal region with a toggle with five layers, two of which correspond to markers of important buildings and parks of the region (R.7), and the remaining represent spatial analyses.

The map view allows the user to see the results of the hotspots and coldspots of the indicators from Ranking 1 (R.5) by plotting red hexagons on top of the map for hotspots and blue hexagons for coldspots. Also, the user can select the layer of the relationship map to visualize the relationship between the scores of Ranking 0 and the sum of the indicator values of Ranking 1 (R.3). For this layer, each hexagon will have assigned a single colour from a four-colour scheme representing the four possible relationships: High-High, High-Low, Low-High, and Low-Low. The user can select and overlap as many layers as wanted. Their overlapping helps the user find relationships, patterns, and spot areas of interest. All views are updated when the user selects a group of data items on the rankings and clicks 'Show selection in the map'. The corresponding hexagons to those areas are shown on the map (R.4). A new causal network is shown on the Causality view considering only the data of the desired areas as Fig. 3.9 shows. Because the need to add new hexagons to the analysis can come up while observing the selected hexagons data, map view allows clicking more hexagons on the map to keep updating the views with more information.

Causality view

The Causality view visualizes a causal network composed of nodes representing the indicators from Ranking 1, connected by links denoting causal influence (R.6). The causality view is shown in Fig. 3.2 C. The graph is created using the entire data of the municipal region when there are no selected items from the rankings. When a selection of items is made, the causal graph will only consider the data from the areas selected as is shown in Figure 3.9. Multiple visual design decisions were taken for a better and easier understanding of the causal graphs. The colour scale of the graph corresponds with the same one from Ranking 1. An indicator has assigned a specific colour that is shown in both views. The size of the nodes varies depending on how many causal relations the variable provokes. The links' width is associated with the causal value of the causal relation. The wider the link, the bigger the causal value. Interactive techniques were also applied to the causal graph. The user can drag and drop the links to organize the nodes and links in a preferred way. When hovering on top of a node, only the causal relations from that node are shown in the view.

Chapter 4

Results and Evaluation

This chapter presents and explains each step of the evaluation of UrbViz. The evaluation was divided into three (3) parts: pre-study questionnaire, execution of tasks using the system with real scenarios, and post-study questionnaire. We invited two data analysts from Halifax Regional Municipality who make urban decisions regarding the datasets included in the tool to perform the user evaluation. Before taking part in it, we provided them with a brief explanation of 10 minutes about the different features and views of the system. The experts did not interact with the tool before this instance. Their first interaction was during the evaluation.

4.1 Pre-study questionnaire

The pre-study questionnaire consisted of seven questions about the current tools the experts were using for making urban decisions. We gathered information about how long they have been using them, their level of helpfulness and satisfaction, the type of issues they face, if the tool provided data visualizations and spatial or statistical analyses, and, finally, if they would change anything of it. This step showed that experts E1 and E2 used Excel, ArcGIS Pro, and GIS for data analysis to make decisions. They both expressed that they were very satisfied, and the tools helped do their analyses. E1 described that the tools provided charts, graphs, and maps, while E2 mentioned that visualization and basic statistical analyses are possible manually. Regarding changes they would like in their current methods, E2 mentioned the helpfulness of generating automatic visualizations.

4.2 Execution of tasks

After the pre-study, the users were given eight tasks involving real usage scenarios while exploring and interacting with the system. The tasks were composed of instructions to simulate a scenario and a final question to assess the conclusions they made from the visualizations shown in it.

Relationship between rankings. The experts were asked to set the scenario shown in Fig. 3.2 A, setting *Residential instability scores* for Ranking 0 and *Total distracted driving*, *Total intersection related*, *Total impaired driving* for Ranking 1 (R.1), and were asked if they found any correlation between the two rankings (R.2.). Both experts could perform this task successfully and made conclusions similar to our expectations: areas with high scores for Ranking 0 are also areas with high values for Ranking 1. E1 stated "*Higher instability leads to a higher likelihood of issues*" and E2 "*Areas with high residential instability score usually have more traffic collisions labelled as Total impaired driving.*" We could verify that the users can successfully interact, play with the rankings, and find relationships between them.

Correlation between hotspots/coldspots and parks/buildings. With the same scenario in Fig. 3.2 A, the experts had to overlap the layers of hotspots of Ranking 1 indicators (R.5) with parks and buildings (R.7), shown in Fig. 3.2 B, to analyze if there was any correlation between the hotspot areas and the location of these entities. After that, they had to repeat the same task with the coldspots. The experts performed these tasks, and both of them could conclude that hotspot areas of *Total distracted driving*, *Total intersection related*, *Total impaired driving* are hexagon areas where a high number of parks and buildings are located and that coldspots areas of the same indicators are areas where no parks and buildings are located.

Major cause event among indicators. For this task, the experts had to add *Total aggressive driving* to Ranking 1 to the scenario in Fig. 3.2 A, explore the causal graph created from the indicators of Ranking 1, and identify the event that majorly causes other events (R.6) shown in Fig. 4.1. E2 could achieve this task

by answering *Total distracted driving*, while E1 answered incorrectly *Total aggressive driving*.

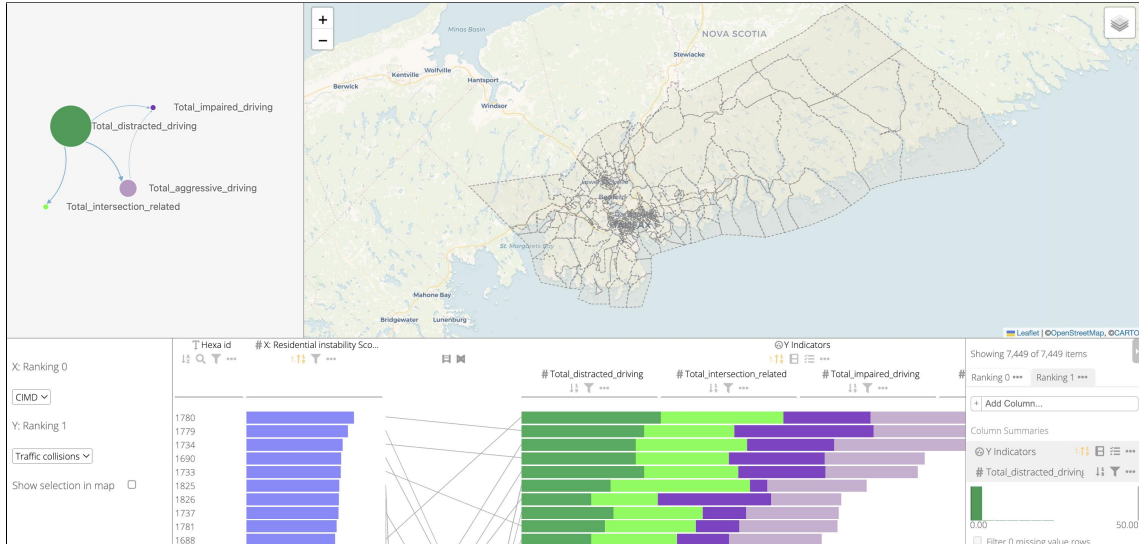


Figure 4.1: Causal graph for the indicators of Ranking 1 for the entire municipality region.

Relationship between index scores and dataset attributes. In this task, the experts had to work with the neighbour claim calls dataset. They had to select *Economic dependency scores* for Ranking 0 and *ROW Maintenance, ROW lights, Traffic Inquiries* for Ranking 1, then display the relationship map layer, shown in Fig. 4.2 (Left), to observe the hexagon areas of a chosen region and analyze if there was any relationship between the index score from Ranking 0 and the dataset attributes from Ranking 1 (R.3). After that, the users had to change the column of Ranking 0 for *Residential instability scores* and analyze the new relationship map, shown in Fig. 4.2 (Right), of the same region. For these tasks, we expected the users to note that areas with a high concentration of people usually have low Economic Dependency scores and high Residential Instability scores, and vice versa for areas with a low concentration of people.

E1 solved this task successfully, stating *"Rural areas appear to have a more 'high-high'"* for the first score, and *"The High high has moved into the urban core, whereas in the economic one, it was more based on rural vs. urban. This is broader across the municipality"* for the second score. E2 first expressed *"There doesn't appear to*



Figure 4.2: Fig. (Left) shows the relationship map between 'Economic dependency scores' and 'ROW Maintenance, ROW lights, Traffic Inquiries'. Fig. (Right) shows the relationship map between 'Residential instability scores' and 'ROW Maintenance, ROW lights, Traffic Inquiries'

be much relationship – all of (region selected) appears to have a High 'Y', while only one small area has a High 'X'." and then *"This time, almost all of (region selected) has a High 'X', instead of just one area."* After these tasks, we wanted to test if the users could interact with multiple spatial layers simultaneously. For this, the users had to select and overlap the layers of hotspots and relationship maps of the scenario previously described and identify which hexagons from the relationship map were coinciding in the hotspots. Both experts answered successfully that most high-high hexagons were overlapping with the hotspots. These results show that the users can explore the map using the layer of the relationship map and create conclusions from it.

Analysis of causal relations among indicators. For this final task, the users had to work with the fire incidents dataset. They had to select the columns of *Residential Instability Scores* for Ranking 0 and *False Alarms, Alarms, Outside Fire* for Ranking 1, then observe the causality graph of the entire region for these 3 indicators, particularly focuses on the causal relation of *False Alarms, Outside Fire* (R.6). The scenario for this task is shown in Figure 4.3. After that, they had to sort descending Ranking 1 by *False Alarms* and visualize the causal graph of the first 10 rows of it

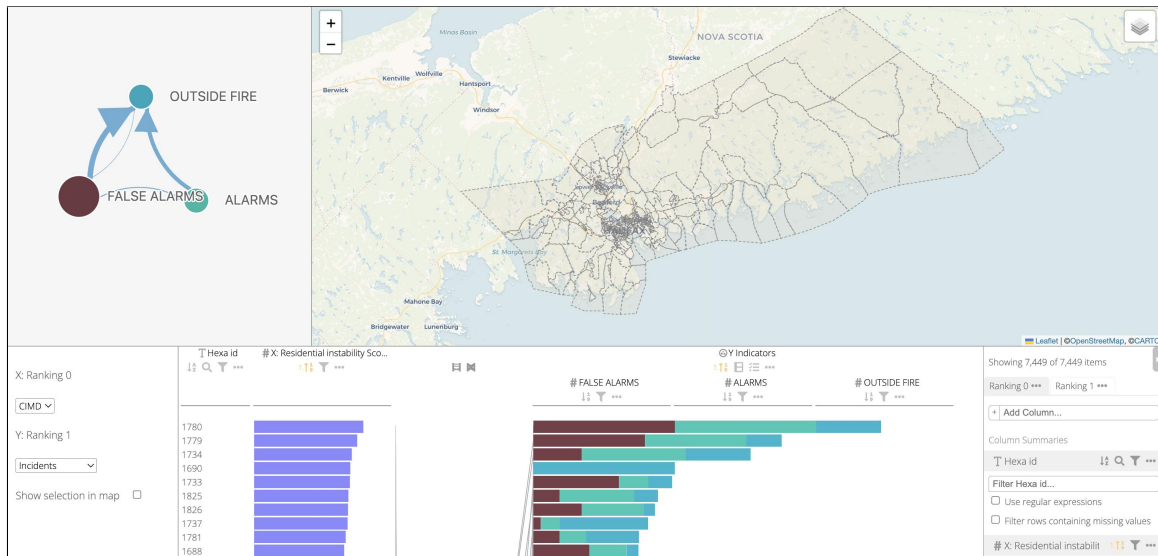


Figure 4.3: Fig. shows the causal graph of the entire municipality region

(R.4) shown in Figure 4.4. We asked the experts if they found any difference between both graphs and their thoughts about why this happened. We were expecting results stating the causal relation switched, and the causal graph of specific areas, especially those with high false alarms, had different causal relations than the entire region. Only E2 succeeded in this task, stating *"The causality arrow switched directions – for the entire region, 'False Alarms' was the cause of 'Outside Fire', but in the areas with the highest numbers of false alarms, 'Outside Fires' causes 'False Alarms'. Perhaps after an actual fire event, people in the area are more cautious and report false alarms"*. We believe E1 did not understand the causality view and could not perform this task. These results give us an insight that only when the experts comprehend the feature of causal graphs they can elaborate successful conclusions from it.

4.3 Post-study questionnaire

The post-study questionnaire consisted of fifteen questions to gather feedback on the usefulness and usability of the tool. The feedback is summarized as follows.

Interpretability and Usefulness Both experts agreed that the system helped them to discover behaviours or patterns in the urban data that they did not know before and were able to draw conclusions or tell stories from all the data presented.

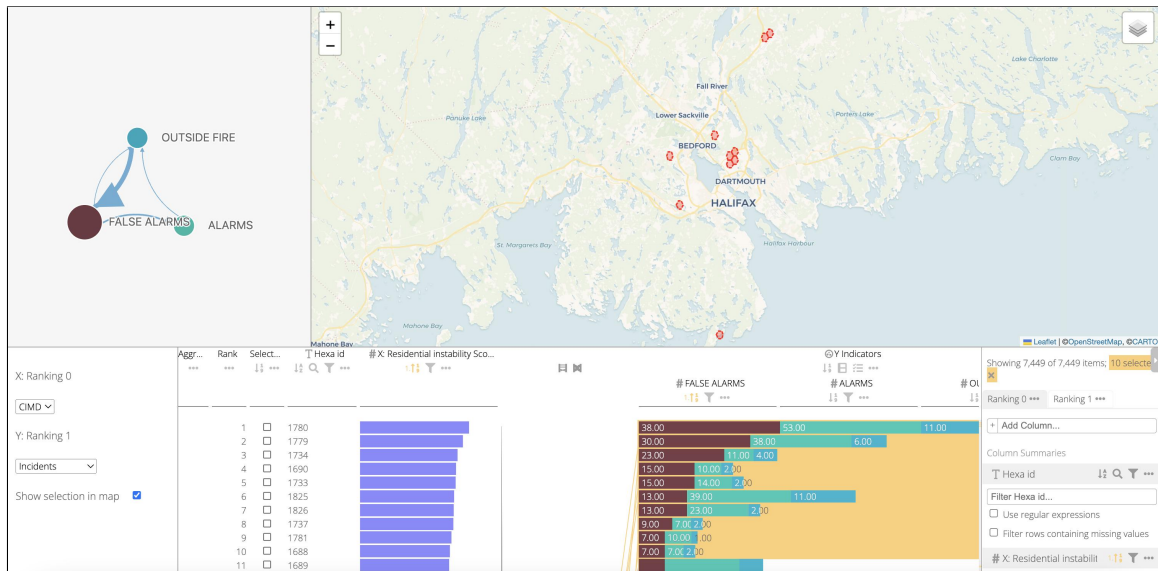


Figure 4.4: Fig. shows the causal graph of the first 10 areas of Ranking 1 sorted by False Alarms.

In particular, E1 stated that the conclusions drawn from the tasks were sensible and reflected actual patterns of the city. Regarding usefulness, both experts stated that the system would help them to make faster and better urban decisions supported by data and that it would be more helpful if they could add other datasets to the system. Both mentioned they would like to use UrbViz on a regular basis.

Usability The experts were neutral about the system being easy to use. In particular, E1 disagreed that people can learn to use the system quickly. However, the experts stated that after the given tutorial, they felt able to investigate other datasets on their own. E1 remarked *"This is a very interesting tool that allows you to compare and summarize quickly. It would be important to have some kind of training before use because it isn't very intuitive at first glance (but makes sense when explained)"*. Similarly, E2 mentioned that, after the initial explanation, they would not need the support of a technical person to be able to use the system and commented *"Overall on a usability end the tool is great if you're used to using tools like this, I think there could be some issues for folks who do not have a comfort level using this kind of tool"*.

4.4 Conclusion

The evaluation of UrbViz performed with urban data analysts from Halifax Regional Municipality completing real usage scenarios provided us with positive feedback in terms of usefulness and usability. The domain experts, after a brief introduction to the system, could interact with the tool, perform the tasks and create conclusions from the information provided by it. They agree that the results they could find were sensible and reflected actual patterns of the city, and they were able to discover new outcomes by exploring the tool. They believe UrbViz would help them to take faster and better urban decisions in their work, and they showed interest in adding new datasets to the tool. The experts mentioned the importance of having a proper introduction to the tool before its use, they do not think a person can quickly learn this tool without an introduction. However, once the user has this knowledge the experts agreed they can use the tool on their own and the support of a technical person is not needed.

Chapter 5

Conclusions and Future Work

This thesis presents a novel visual analytics approach that assists urban data analysts in identifying patterns and relationships among indices and multiple urban datasets. Working closely with domain experts from Halifax Regional Municipality, we identified a specific problem in their data analysis process and proposed a system comprised of multi-attribute rankings, spatial analyses, and geo-located causal analysis.

During this journey, we faced different issues in the interaction with the municipality. In the first stage of data collection, we had a long wait until we gathered the data due to privacy terms. Later, we found out each dataset belonged to a different department in the municipality, leading each team to have its own convention for data storage. For instance, teams used a different coordinate projection to register the events. Because of this, the step of data preprocessing took a longer period of time than expected, since we had to set a convention for all datasets and make the proper modifications.

We then develop UrbViz by coordinating a set of well-designed visualizations. During this stage, we faced the challenge of balancing both the creation of a novel combination of visualizations and the provision of a user-friendly visualization system that could be used by urban analysts from a municipality.

Finally, we conducted a user evaluation of real-world datasets by two domain experts. For this last stage, we conducted two in-person meetings in the Halifax Regional Municipality. In the first one, we introduced and gave a live demonstration of UrbViz to 10 employees. The second meeting was a user evaluation session intended for the people who had taken part in the initial demonstration to use the tool for the first time in a supervised setting. Due to complications, none of the people from the first meeting were available. This meant we had to conduct the user evaluation with two people unfamiliar with the tool. Despite the fact that these people were unfamiliar with the tool the evaluation was a success providing us with positive feedback about

its usefulness and usability.

5.1 Limitations and Future Work

Three limitations are observed in this study. First, the system does not support the option of filtering the analyses by time. The computations for each feature in UrbViz take into consideration the entire time frame of the data collection, which makes the calculations not very precise. In the future, we will include a timeline feature in which the user can select the time frame desired and the analyses will be calculated by taking only the observations that happened in that lapse. Secondly, the system supports the comparison between an index and a custom multi-attribute ranking, but the option of comparing different custom rankings is not allowed. In the future, we want to improve the system by adding the possibility of saving rankings and allowing the users to make comparisons between them. Finally, the time calculation for the spatial analyses exceeds fifteen seconds. In the future, we will optimize the functions, implement them with a higher-performance programming language, and deploy it on a distributed computing server.

Bibliography

- [1] Volodymyr Agafonkin. Leaflet.js.
- [2] L. Anselin. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht, 1988.
- [3] Luc Anselin. A local indicator of multivariate spatial association: Extending geary's c. *Geographical Analysis*, 51(2):133–150, 2019.
- [4] Luc Anselin. *Geoda: An introduction to spatial data science*, 2020.
- [5] Luc Anselin and Oleg O. Smirnov. Efficient algorithms for constructing proper higher order spatial lag operators. *Journal of Regional Science*, 36:67–89, 1996.
- [6] Grace Antony and K Rao. A composite index to explain variations in poverty, health, nutritional status and standard of living: Use of multivariate statistical methods. *Public health*, 121:578–87, 09 2007.
- [7] Mike Bostock, Jason Davies, Jeffrey Heer, Vadim Ogievetsky, and community. D3.js data-driven documents.
- [8] Aoxiang Cao, Yuanyuan Qiao, Kewu Sun, Hao Zhang, and Jie Yang. Network traffic analysis and prediction of hotspot in cellular network. *2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, pages 452–456, 2018.
- [9] Wei Chen, Zhaosong Huang, Feiran Wu, Minfeng Zhu, Huihua Guan, and Ross Maciejewski. Vaud: A visual analysis approach for exploring spatio-temporal urban data. *IEEE Transactions on Visualization and Computer Graphics*, 24(9):2636–2648, September 2018. Funding Information: This research has been sponsored in part by the National 973 Program of China (2015CB352503), Major Program of National Natural Science Foundation of China (61232012), National Natural Science Foundation of China (61422211, 61772456), and National Natural Science Foundation of China (U1609217). Publisher Copyright: © 1995-2012 IEEE.
- [10] Lawrence J. Christiano. Christopher a. sims and vector autoregressions*. *The Scandinavian Journal of Economics*, 114(4):1082–1104, 2012.
- [11] Laura García Cuenca, Nourdine Aliane, Enrique Puertas, and Javier Fernandez Andres. Traffic hotspots visualization and warning system. In *2018 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, pages 1–5, 2018.

- [12] Jose Florencio de Queiroz Neto, Emanuele Santos, Creto Vidal, and David Ebert. A visual analytics approach to facilitate crime hotspot analysis. *Computer Graphics Forum*, 39:139–151, 06 2020.
- [13] Zikun Deng, Di Weng, Xiao Xie, Jie Bao, Yu Zheng, Mingliang Xu, Wei Chen, and Yingcai Wu. Compass: Towards better causal analysis of urban time series. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 10 2021.
- [14] Cecilia di Sciascio, Vedran Sabol, and Eduardo Veas. Rank as you go: User-driven exploration of search results. *Proceedings of the 21st International Conference on Intelligent User Interfaces*, page 118–129, 2016.
- [15] D. Dickey and Wayne Fuller. Distribution of the estimators for autoregressive time series with a unit root. *JASA. Journal of the American Statistical Association*, 74, 06 1979.
- [16] Richard Dunn. A dynamic approach to two-variable color mapping. *The American Statistician*, 43(4):245–252, 1989.
- [17] J. DURBIN and G. S. WATSON. TESTING FOR SERIAL CORRELATION IN LEAST SQUARES REGRESSION. II. *Biometrika*, 38(1-2):159–178, 06 1951.
- [18] Germain Garcia-Zanabria, Erick Gomez-Nieto, Jaqueline Silveira, Jorge Poco, Marcelo Nery, Sergio Adorno, and Luis G. Nonato. Mirante: A visualization tool for analyzing urban crimes. In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 148–155, 2020.
- [19] Germain García-Zanabria, Marcos M. Raimundo, Jorge Poco, Marcelo Batista Nery, Cláudio T. Silva, Sergio Adorno, and Luis Gustavo Nonato. Cripav: Street-level crime patterns analysis and visualization. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4000–4015, 2022.
- [20] Germain Garcia, Jaqueline Silveira, Jorge Poco, Afonso Paiva, Marcelo Batista Nery, Claudio T. Silva, Sérgio Adorno, and Luis Gustavo Nonato. Crimanalyzer: Understanding crime patterns in são paulo. *IEEE Transactions on Visualization and Computer Graphics*, 27(4):2313–2328, 2021.
- [21] Alex Godwin and John T. Stasko. Hotsketch: Drawing police patrol routes among spatiotemporal crime hotspots. In Tung Bui, editor, *50th Hawaii International Conference on System Sciences, HICSS 2017, Hilton Waikoloa Village, Hawaii, USA, January 4-7, 2017*, pages 1–9. ScholarSpace / AIS Electronic Library (AISeL), 2017.
- [22] C W J Granger. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, 37(3):424–438, July 1969.

- [23] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19:2277–86, 12 2013.
- [24] Rich Harris. Svelte.js.
- [25] Xiaoke Huang, ye Zhao, Chao Ma, Jing Yang, Xinyue ye, and Chong Zhang. Trajgraph: A graph-based visual analytics approach to studying. *IEEE Transactions on Visualization and Computer Graphics*, 22:1–1, 1 2015.
- [26] Robin John Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Australia, 2nd edition, 2018.
- [27] Quan Li, Q. Liu, C. Tang, Z. Li, S. Wei, X. Peng, M. Zheng, Tianjian Chen, and Q. Yang. Warehouse vis: A visual analytics approach to facilitating warehouse location selection for business districts. *Computer Graphics Forum*, 39:483–495, 06 2020.
- [28] Dongyu Liu, Di Weng, Yuhong li, Jie Bao, Yu Zheng, Huamin Qu, and Yingcai Wu. Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE Transactions on Visualization and Computer Graphics*, 23:1–1, 01 2016.
- [29] Yafeng Lu, Hong Wang, Steven Landis, and Ross Maciejewski. A visual analytics framework for identifying topic drivers in media events. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 09 2017.
- [30] Fabio Miranda, Harish Doraiswamy, Marcos Lage, Kai Zhao, Bruno Gonçalves, Luc Wilson, Mondrian Hsieh, and Claudio Silva. Urban pulse: Capturing the rhythm of cities. *IEEE Transactions on Visualization and Computer Graphics*, 23:1–1, 01 2016.
- [31] Patrick Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37:17–23, 07 1950.
- [32] Stephan Pajer, Marc Streit, Thomas Torsney-Weir, Florian Spechtenhauser, Torsten Möller, and Harald Piringer. Weightlifter: Visual weight space exploration for multi-criteria decision making. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):611–620, 2017.
- [33] ArcGIS Pro. How hot spot analysis (getis-ord gi*) works, 2015.
- [34] ArcGIS Pro. How spatial autocorrelation (global moran’s i) works, 2015.
- [35] C. H. S. Santos, Maryam Hosseini, Joao Rulff, Nivan Ferreira, Luc Wilson, Fábio Miranda, Cláudio T. Silva, and Marcos Lage. A visual analytics system for profiling urban land use evolution. *ArXiv*, abs/2112.06122, 2021.

- [36] Andreas Scheidl, Roger A. Leite, and Silvia Miksch. VisMiFlow: Visual Analytics to Support Citizen Migration Understanding Over Time and Space. In Marco Agus, Christoph Garth, and Andreas Kerren, editors, *EuroVis 2021 - Short Papers*. The Eurographics Association, 2021.
- [37] Antonio Scipioni, Anna Mazzi, Marco Mason, and Alessandro Manzardo. The dashboard of sustainability to measure the local urban sustainable development: The case study of padua municipality. *Ecological Indicators*, 9(2):364–380, 2009.
- [38] Daniel Seebacher, Matthias Miller, Thomas Polk, Johannes Fuchs, and Daniel Keim. Visual analytics of volunteered geographic information: Detection and investigation of urban heat islands. *IEEE Computer Graphics and Applications*, PP:1–1, 07 2019.
- [39] Dani Arribas-Bel Sergio J. Rey and Levi J. Wolf. Geographic data science with python, 2020.
- [40] Lin Shao, Sebastian Mittelstädt, Ran Goldblatt, Itzhak Omer, Peter Bak, and Tobias Schreck. Streetexplorer: Visual exploration of feature-based patterns in urban street networks. In Nadia Magnenat-Thalmann, Paul Richard, Lars Linsen, Alexandru Telea, Sebastiano Battiato, Francisco H. Imai, and José Braz, editors, *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 2: IVAPP, Rome, Italy, February 27-29, 2016*, pages 86–97. SciTePress, 2016.
- [41] Priya Singh, Chandra Sabnani, and Vijay Kapse. Hotspot analysis of structure fires in urban agglomeration: A case of nagpur city, india. *Fire*, 4:38, 07 2021.
- [42] João Soares, Maria Marquês, and Carlos Monteiro. A multivariate methodology to uncover regional disparities: A contribution to improve european union and governmental decisions. *European Journal of Operational Research*, 145:121–135, 02 2003.
- [43] StatsCan. The canadian index of multiple deprivation, 2016.
- [44] Paolo Tasserone, Hestia Zinsmeister, Liselotte Rambonnet, Auke-Florian Hiemstra, Daniël Siepman, and Tim van Emmerik. Plastic hotspot mapping in urban water systems. *Geosciences*, 10:342, 08 2020.
- [45] Chun-Hua Tsai and Peter Brusilovsky. Beyond the ranked list: User-driven exploration and diversification of social recommendation. In *23rd International Conference on Intelligent User Interfaces*, IUI '18, page 239–250, New York, NY, USA, 2018. Association for Computing Machinery.
- [46] Dong-Bach Vo, Kristina Lazarova, Helen C. Purchase, and Mark McCann. Visual causality: Investigating graph layouts for understanding causal processes.

- In Ahti-Veikko Pietarinen, Peter Chapman, Leonie Bosveld-de Smet, Valeria Giardino, James Corter, and Sven Linker, editors, *Diagrammatic Representation and Inference*, pages 332–347, Cham, 2020. Springer International Publishing.
- [47] Emily Wall, Subhajit Das, Ravish Chawla, Bharath Kalidindi, Eli Brown, and Alex Endert. Podium: Ranking data using mixed-initiative visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 08 2017.
- [48] Jun Wang and Klaus Mueller. The visual causality analyst: An interactive interface for causal reasoning. *IEEE Transactions on Visualization and Computer Graphics*, 22:1–1, 11 2015.
- [49] Jun Wang and Klaus Mueller. Visual causality analysis made practical. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 151–161, 2017.
- [50] Di Weng, Ran Chen, Zikun Deng, Feiran Wu, Jingmin Chen, and Yingcai Wu. Srvis: Towards better spatial integration in ranking visualization. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 09 2018.
- [51] Zhijie Xiao. Testing the null hypothesis of stationarity against an autoregressive unit root alternative. *Journal of Time Series Analysis*, 22:87 – 105, 12 2001.
- [52] Xiao Xie, Fan Du, and Yingcai Wu. A visual analytics approach for exploratory causal analysis: Exploration, validation, and applications. *IEEE transactions on visualization and computer graphics*, PP, 10 2020.
- [53] Xiao Xie, Moqi He, and Yingcai Wu. Causalflow: Visual analytics of causality in event sequences. *ArXiv*, abs/2008.11899, 2020.

Appendix A

Letter of Approval

The following pages include a copy of the Letter of Approval from the Social Sciences & Humanities Research Ethics Board.

**Social Sciences & Humanities Research Ethics Board
Letter of Approval**

October 05, 2022
Micaela Melo
Computer Science\Computer Science

Dear Micaela,

REB #: 2022-6228
Project Title: A Visual Analytics System for the Analysis of City Indicators based on Multivariate Indices

Effective Date: October 05, 2022
Expiry Date: October 05, 2023

The Social Sciences & Humanities Research Ethics Board has reviewed your application for research involving humans and found the proposed research to be in accordance with the Tri-Council Policy Statement on *Ethical Conduct for Research Involving Humans*. This approval will be in effect for 12 months as indicated above. This approval is subject to the conditions listed below which constitute your on-going responsibilities with respect to the ethical conduct of this research.

Sincerely,

Dr. Megan Bailey
Chair, Social Sciences and Humanities Research Ethics Board
Dalhousie University

Funded: Mitacs accelerate.
Partner organization: Halifax Regional Municipality
IT23586

Post REB Approval: On-going Responsibilities of Researchers

After receiving ethical approval for the conduct of research involving humans, there are several ongoing responsibilities that researchers must meet to remain in compliance with University and Tri-Council policies.

1. Additional Research Ethics approval

Prior to conducting any research, researchers must ensure that all required research ethics approvals are secured (in addition to Dalhousie approval). This includes, but is

not limited to, securing appropriate research ethics approvals from: other institutions with whom the PI is affiliated; the institutions of research team members; the institution at which participants may be recruited or from which data may be collected; organizations or groups (e.g. school boards, Indigenous communities, correctional services, long-term care facilities, service agencies and community groups) and from any other responsible review body or bodies at the research site.

2. Reporting adverse events

Any significant adverse events experienced by research participants must be reported **in writing** to Research Ethics **within 24 hours** of their occurrence. Examples of what might be considered “significant” include: a negative physical reaction by a participant (e.g. fainting, nausea, unexpected pain, allergic reaction), an emotional breakdown of a participant during an interview, report by a participant of some sort of negative repercussion from their participation (e.g. reaction of spouse or employer) or complaint by a participant with respect to their participation, report of neglect or abuse of a child or adult in need of protection, or a privacy breach. The above list is indicative but not all-inclusive. The written report must include details of the situation and actions taken (or proposed) by the researcher in response to the incident.

3. Seeking approval for changes to research

Prior to implementing any changes to your research plan, whether to the risk assessment, methods, analysis, study instruments or recruitment/consent material, researchers must submit them to the Research Ethics Board for review and approval. This is done by completing the amendment request process (described on the website) and submitting an updated ethics submission that includes and explains the proposed changes. Please note that reviews are not conducted in August.

4. Continuing ethical review - annual reports

Research involving humans is subject to continuing REB review and oversight. REB approvals are valid for up to 12 months at a time (per the Tri-Council Policy Statement (TCPS) article 6.14). Prior to the REB approval expiry date, researchers may apply to extend REB approval by completing an Annual Report (available on the website). The report should be submitted 3 weeks in advance of the REB approval expiry date to allow time for REB review and to prevent a lapse of ethics approval for the research. Researchers should note that no research involving humans may be conducted in the absence of a valid ethical approval and that allowing REB approval to lapse is a violation of the University Scholarly Misconduct Policy, inconsistent with the TCPS and may result in the suspension of research and research funding, as required by the funding agency.

5. Final review - final reports

When the researcher is confident that all research-related interventions or interactions with participants have been completed (for prospective research) and/or that all data acquisition is

complete, there will be no further access to participant records or collection of biological materials (for secondary use of information research), a Final Report (available on the website) must be submitted to Research Ethics. After review and acknowledgement of the Final Report, the Research Ethics file will be closed.

6. Retaining records in a secure manner

Researchers must ensure that records and data associated with their research are managed consistent with their approved research plans both during and after the project. Research information must be confidentially and securely retained and/or disposed of in such a manner as to comply with confidentiality provisions specified in the protocol and consent forms. This may involve destruction of the records, or continued arrangements for secure storage.

It is the researcher's responsibility to keep a copy of the REB approval letters. This can be important to demonstrate that research was undertaken with Board approval. Please note that the University will securely store your REB project file for 5 years after the REB approval end date at which point the file records may be permanently destroyed.

7. Current contact information and university affiliation

The lead researchers must inform the Research Ethics office of any changes to contact information for the PI (and supervisor, if appropriate), especially the electronic mail address, for the duration of the REB approval. The PI must inform Research Ethics if there is a termination or interruption of his or her affiliation with Dalhousie University.

8. Legal Counsel

The Principal Investigator agrees to comply with all legislative and regulatory requirements that apply to the project. The Principal Investigator agrees to notify the University Legal Counsel office in the event that he or she receives a notice of non-compliance, complaint or other proceeding relating to such requirements.

9. Supervision of students

Faculty must ensure that students conducting research under their supervision are aware of their responsibilities as described above and have adequate support to conduct their research in a safe and ethical manner.